THE SUBNUCLEAR LOCALISATION OF NOTCH RESPONSIVE GENES

Matthew Jones

Jesus College

Supervisor: Sarah Bray

Department of Physiology, Development and Neuroscience

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

March 2018

The subnuclear localisation of Notch responsive genes – Matthew Jones

March 2018

For Mum, Dad, Gaby and Dan.

“Time’s fun when you’re having flies.”

DECLARATION

The work described in this thesis was carried out in the laboratory of Professor Sarah Bray in the department of Physiology, Development and Neuroscience within the University of Cambridge, between September 2013 and June 2017.

This dissertation is the result of my own work and includes nothing, which is the outcome of work done in collaboration, except where specifically indicated in the acknowledgement and in the text.

Neither this part, nor any part of it has been previously submitted for any degree or qualification at this, or any other university.

This thesis does not exceed the word limit set by the Degree committee of Biology.

Signed: 

Matthew Jones

BA (Hons), MPhil, MA (Cantab).

Cambridge, UK

28th September 2017

ABSTRACT

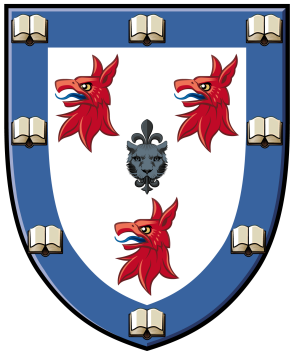
Title: The subnuclear localisation of Notch responsive genes.

Candidate Name: Matthew Jones

Notch signalling is a highly conserved cell-cell communication pathway with critical roles in metazoan development and mutations in Notch pathway components are implicated in many types of cancer. Notch is an excellent and well-studied model of biological signalling and gene regulation, with a single intracellular messenger, one receptor and two ligands in *Drosophila*. However, despite the limited number of chemical players involved, a striking number of different outcomes arise. Molecular studies have shown that Notch activates different targets in different cell types and it is well known that Notch is important for maintaining a stem cell fate in some situations and driving differentiation in others. Thus some of the factors affecting the regulation of Notch target genes are yet to be discovered.

Previous studies in various organisms have found that the location of a gene within the nucleus is important for its regulation and genome reorganisation can occur following gene activation or during development. Therefore this project aimed to label individual Notch responsive loci and determine their subnuclear localisation. In order to tag loci of interest a CRISPR/Cas9 genome-editing method was established that enabled the insertion of locus tags at Notch targets, namely the well-characterized *Enhancer of split* locus and also *dpn* and *Hey*, two transcription factors involved in neural cell fate decisions.

The ParB/Int system is a recently developed locus tagging system and is not well characterised in *Drosophila.* It has a number of advantages over the traditional LacO/LacI-GFP locus tagging system as it does not rely on binding site repeats for signal amplification and can label two loci simultaneously in different colours. This thesis characterised the ParB/Int system in the *Drosophila* salivary gland and larval L3 neuroblast. Using 3D image segmentation hundreds of nuclei were reconstructed and a volume based normalisation method was applied to determine the subnuclear localisation of several Notch targets with and without genetic manipulations of the Notch pathway.



ACKNOWLEDGEMENTS

First of all I would like to thank my supervisor Sarah for providing an exciting PhD project, encouraging me to pursue new technologies and persevering through my learning curves. I am very grateful for all the time invested in giving me feedback on my lab meetings, presentations and reports. I would also like to thank Sarah for the many lab retreats in pleasant corners of the U.K. and Spain, which were hard work but ultimately rewarding and a lot of fun. I owe innumerable thanks to Kat Millen, who not only ran the lab impeccably throughout my time here but also performed microinjection for all of the transgenesis required for my project and provided guidance and moral support on many occasions. Thanks also to Agnès for being a delightful colleague to work with, keeping our flies alive in collaboration with Kat, and brightening the world with her artwork. I am deeply indebted to Gustavo Cerda-Moya for answering countless molecular biology questions, training me and supporting me through experimental failures. Thanks to Robert Stojnic, Radu Zabet and Bettina Fischer for being patient tutors in R programming and computational biology. Thank you to Jinhua Li and Lenka Skalska for molecular biology training too. I'm very grateful to Maria Gomez-Lamarca for supporting me in the realms of fly genetics and teaching me a great deal about imaging. A huge thanks to Eva Zacharioudaki for being a fountain of knowledge about neuroblasts, flywork, imaging, quantification and for kindly preparing neuroblast cultures allowing me to image the locus tags live. Thank you to Hadi Boukhatmi, Sara Morais da Silva, and Shiyun Feng for being helpful and supportive colleagues and to Silvie Fexova, Zoe Pillidge and Julia Falo-Sanjuan for being fun colleagues and sharing lots of useful knowledge. A special mention to Stella 'Lymphidaki' Lempidaki and Stephen Chan for so much moral support, experimental advice and many fun social times. Thanks also to the Genes and Morphology group and all the members of P.D.N. who supported me along the way. I owe a special thanks to Rob White for being an excellent advisor on troubleshooting, reports, project management and keeping me focused on the positives. I cannot thank you enough.

Thank you to my good friends in Cambridge, and now beyond in some cases, Tara Finegan, Rob Tetley, Helen Fox, Hannah Green, Joy Thompson, Jonathan Lawson, Vasja Urbancic, Petros Giannaros, Velislava Petrova, John Lees, Matthew Cornwell, Natalia Rossetti, Raphael Hetherington, Henry Staples, Will Lewis-Smith, Emma

Robson, Ed Blows, Marino Krstulović and Max Goulding. I am grateful to all the friends I made at Jesus College for the many fun times and huge amount I have gained from being a part of the Jesus MCR for 5 years. I am especially grateful to Shailaja Fennell for being a fantastic Deputy Graduate Tutor.

Thank you to Jim Haseloff, Raul Duran and Jeremy Minshull for supporting me through exciting summer projects that prepared me for my PhD. Thank you to Paul Elliott at Homerton College, Michael Bate at King's College, and Matthew Oswald in the Landgraf lab without whom I would not have got started in *Drosophila* research. Many, many thanks to Matthias Landgraf for encouraging me first as my Part II supervisor, inspiring me to embark upon a PhD and for supporting me along the way.

Special thanks to Joseph Hanly and Marcia Kishida for being fantastic friends through the entire rollercoaster.

I would like to thank all my friends and family members who supported me throughout this endeavour. My family back home, Grandma Merle, Auntie Diane, Shaun and Michael, Auntie Maria, Marc and Sarah and also to Suzanne and Stephen Clennell and Margaret Bell who have been kind, generous and supportive friends for my whole life. Rest in peace Grandad Ken, Grandad Peter and Laurie Bell, you are not forgotten.

Much love and thanks to Irina, draga mea, thanks so much for your sound advice, love and support.

I would not be here without the love and support of my father Les, mother Christine, sister Gabriella and brother Daniel. Thanks so much to all of you and for everything you do to support me.

Table of Contents

LIST OF FIGURES

List of Abbreviations and Acronyms

2D –two-dimensional

3C – chromosome conformation capture

3D – three-dimensional

86Fb – cytogenetic identifier for a control locus tag location (Bischof et al., 2007)

ADAM10 – a metalloendopeptidase that cleaves the Notch receptor during transduction

AmpR – Ampicillin resistance cassette

ANOVA – analysis of variance

ASH2 – a histone methyltransferase subunit

attB/attP – a ΦC31 integrase recombination site

attL/attR – sequence element resulting from ΦC31 integrase recombination

BBT – A buffer made with phosphate buffered saline, bovine serum albumin and Triton

BFM – Bearded Family Member – a group of transcription factors some of which occupy the *Enhancer of split* complex

bHLH – basic helix-loop-helix – DNA binding (b) and dimerization domain (HLH) found in many dimeric transcription factors

Cas9 – CRISPR associated protein 9 – a bacterial RNA programmable nuclease

cDNA – complimentary DNA

ChIP – chromatin immunopreciptiation

CNS – central nervous system (brain lobes and ventral nerve cord)

Cre – Cre recombinase

CRISPR – clustered regularly interspaced palindromic repeats

crRNA – CRISPR RNA

CSL – CBF1/Suppressor of Hairless/Lag-1 – generic name for Notch pathway transcription factor

Ctrl – control

Cy3/Cy5 – fluorescent dyes with red/far red emission spectra

D1 – Delta, a Notch ligand

DamID – DNA adenine methylase identification

ddH_2O – double distilled water

***dpn* – *deadpan,* a Notch target regulating neuroblast proliferation**

Dpn – protein encoded by the *deadpan* gene

dpn*.attP – attP site insertion neighbouring *dpn

***dpn*.Int2 – Int2 insertion at dpn.attP**

DsRed – *Discosoma sp.* Red Express an engineered red fluorescent protein

DsRed.attP – plasmid insert containing DsRed flanked by loxP sites and fused to attP

ECDF – Empirical cumulative distribution frequency

EGFP – enhanced green fluorescent protein

***E(spl)* – Enhancer of split**

E(spl)-mγ – fullnames of the *E(spl)* bHLH proteins were shortened for brevity, full names include the HLH suffix as in the flybase database e.g. E(spl)-mγ-HLH.

EVF – Eroded Volume Fraction

F1 – first familial generation

FA – formaldehyde

FACS – fluorescence activated cell sorting

FIJI – Fiji is just image J

FISH – fluorescence *in situ* hybridisation

FITC – a fluorescent dye with emission in the green range of the visible spectrum

FPKM – fragments per Kilobase per million reads

FRAP – fluorescence recovery after photobleaching

Gal4 – yeast transcriptional activator used to drive transgene expression

GATC – DpnI/DpnII recognition motif

GFP – Green fluorescent protein

GOI – gene of interest

gRNA – guide RNA

H3K27me3 – Histone three, lysine 27 trimethylation

H3K4me2 – Histone three lysine 4 dimethylation

H3K9me3 – Histone three lysine 9 trimethylation

HDR – homology directed repair

***Hey* – Hairy/E(spl)-related with YRPW motif**

***Hey*.Int1 – Int1 insertion adjacent to the *Hey* locus**

Hoechst – Hoechst 33342 DNA stain

Int/Int1/Int2 – ParB recognition sequence

KO – knockout

KS – Kolmogorov-Smirnov test

L3 – third larval stage

LacI – bacterial transcription factor recognising LacO

LacO – sequence motif recognised by LacI

LAD – lamina-associated domain

loxP – Cre recombinase recognition site

Mam – Mastermind – coactivator in the Notch pathway

mCherry – monomeric red fluorescent protein

mRNA – messenger ribonucleic acid

MS2 – a bacteriophage coat protein used to tag transcripts for live imaging

N – the protein encoded by the *Notch* gene

NICD – the intracellular fragment of the Notch receptor protein

NRE – Notch responsive element, a synthetic Notch signalling reporter

Nup – nuclear pore protein/component

ParB/ParB1/ParB2 – two DNA binding proteins from the *B.cenocapcia* ParABS system that recognise parS sequences

parS – alternative name for the Int motifs recognised by ParB proteins

PBS – phosphate buffered saline

PBT– a buffer made with phosphate buffered saline and Triton X-100

pCFD2 – plasmid used to make the nanos-Cas9 fly line

pCFD3 – gRNA expression vector

PCR – polymerase chain reaction

PFA - paraformaldehyde

ΦC31 – bacteriophage from which the attB/attP transgenesis system derives

pInt1/pInt2 – attB containing plasmids bearing parS sequences

qPCR – quantitative PCR

RNA - ribonucleic acid

RNAi – RNA interference

ROI – region of interest

RTK – receptor tyrosine kinase

TAD – Topologically associating domain

TetR/TetO –bacterial repressor protein/binding site pair

tracrRNA – essential component of type II CRISPR systems upon which CRISPR/Cas9 genome editing is based.

U6 – RNA expression promoter used in pCFD3

UAS – Upstream Activator Sequence – yeast enhancer sequence recognised by Gal4

LIST OF APPENDICES

1 Introduction

1.1 Notch signalling

1.1.1 A cell-cell communication pathway

Notch signalling is implicated in a wide array of biological processes such as the regulation of cell proliferation and selection of cell fate. The Notch gene, which encodes the receptor in this pathway, was first identified because heterozygous mutations resulted in small notches on the fly's wing. Now it is evident that this pathway is also a critical regulator of stem cell division and differentiation in the fruit fly central nervous system and the mammalian brain. Furthermore, essentially the same molecular components that maintain the stem cell properties of self-renewal at one stage of development later drive the selection of a terminal cell fate decision. The ability of Notch to take on diverse roles in different contexts is fascinating and raises many pressing questions in developmental biology.

The key components of the Notch pathway have been well characterised by genetic and biochemical studies. More recently the composition of the chromatin with which it interacts is becoming clear through molecular and genomic characterisation of the epigenetic regulation of Notch target genes. In spite of these efforts, the features of Notch gene regulation that allow the same pathway to elicit different responses at

different times in development are not yet fully understood. One possibility is that there are changes in the nuclear organization, as several recent studies suggest that gene position varies depending on activity. To address this we need to know where in the nucleus Notch target-genes reside and whether they are dynamic over time or in response to the signal. The aims of this project were to generate and characterise tools that enable reliable localisation of key Notch target genes and to quantify the nuclear organisation of Notch targets in cell types with and without active Notch signalling, in order to determine whether there is a change in the subnuclear localisation of responsive genes.

1.1.2 Transduction of the Notch signal

Notch signalling is a highly conserved pathway with no signal amplification at the intracellular messenger stage. Metazoans from nematode worms to large mammals all rely on Notch signalling for regulating critical developmental processes and cell fate decisions. These roles can be grouped into three main categories, lateral inhibition, lineage decisions and boundary formation (Bray, 1998). Despite its myriad roles the pathway has a relatively simple transduction step. The signal is transduced via cell-cell contact mediated by a ligand on the sending cell and a receptor on the receiving cell (Bray, 2016). In *Drosophila* there is just one receptor protein, Notch, and two ligands, Delta and Serrate. Contact between ligand and receptor is followed by cleavage of the extracellular fragment of the receptor by the ADAM10 protease and then cleavage of the intracellular domain by γ-secretase. The Notch intracellular domain (NICD) then translocates to the nucleus where it forms a complex with the key transcription factor in the Notch pathway, CSL or Suppressor of Hairless in *Drosophila,* and a co-activator called Mastermind (Figure 1.1). Essentially the same process occurs in mammals but the number of players is expanded, for example there are four receptors Notch1-4, three ligands in the Delta family, and two homologs of Serrate in the Jagged family, giving a total of five ligands (Hori et al., 2013). For the remainder of this thesis, the *Drosophila* nomenclature will be used to refer to the key pathway components.

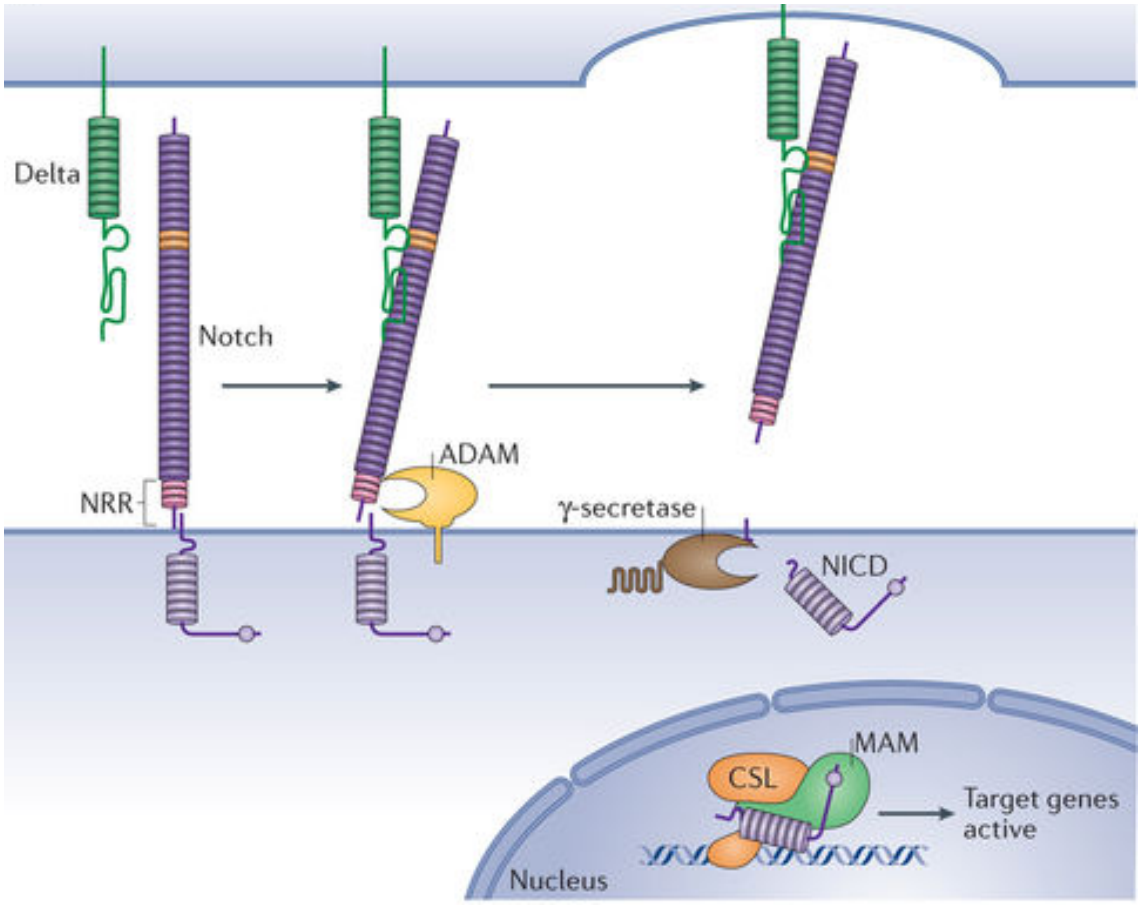


Figure 1.1: Mechanism and key players in canonical Notch signalling from (Bray, 2016).

The core transcription factor is named Suppressor of Hairless (Su(H)) and this component acts as a transcriptional repressor until it forms a complex with the Notch intracellular domain and an important co-activator, Mastermind (Mam) (Bray, 2006). Prior to Notch activation some genes are already bound by Su(H) and co-repressors such as Hairless. This repressor complex is replaced by a Su(H) complex with NICD and Mam, whilst at other loci Notch signalling triggers *de novo* Su(H) binding.

1.1.3 Responses to Notch depend on context

Molecular analysis of the targets of Notch in cell culture lines have found that the cohort of activated genes can differ considerably between cell types (see Figure 1.2) (Krejci et al., 2009). Genetic analysis in flies has revealed that different Notch targets are active in closely related neural cell types, as illustrated by the expression of *deadpan (dpn)* in neuroblasts (San-Juán and Baonza, 2011) and *Hey* (Monastirioti et al., 2010) in the progeny of the neuroblast, destined to be neurons.

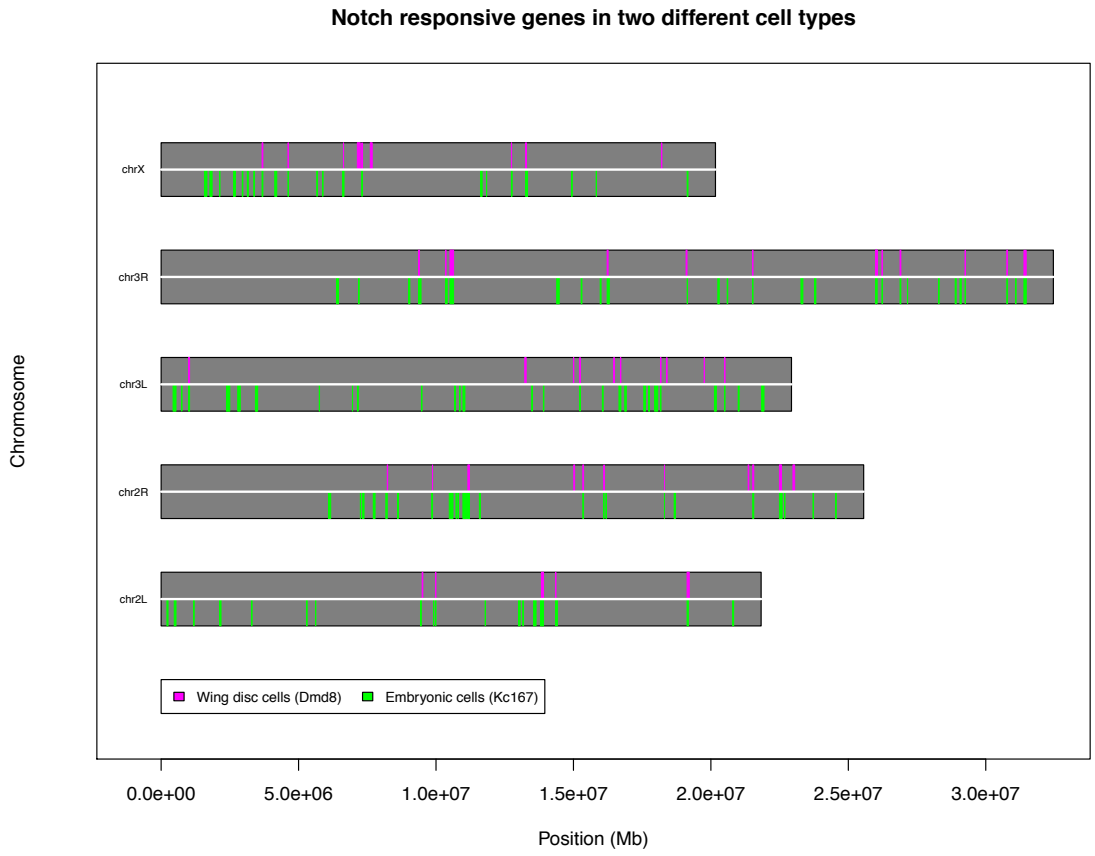


Figure 1.2: Different genes respond to Notch in different cell types.

Expression profiling data from (Krejci et al., 2009) plotted to scale on the five major arms of *Drosophila* chromosomes. Vertical lines represent genes found to be significantly upregulated in the presence of a chemical treatment that activates the Notch pathway (significance level 0.05).

Combined feed-forward and feedback elements of the Notch regulatory network have been identified (Krejci et al., 2009) that could explain some of the context dependency of Notch as they can poise Notch targets to adopt one of several potential outcome states in a short space of time. Moreover, the importance of signalling cross-talk between Notch and RTK signalling has been highlighted by the discovery that many targets are shared between the two (Hurlbut et al., 2009). This indicates that the response to Notch can be modified by the integration of contextual information by multiple inputs acting through common enhancers or by independent enhancers that converge on a single target.

Notch signalling regulates growth through multiple pathways and later, differentiation in the same tissue. The imaginal discs are rapidly growing tissues in the larvae of

Drosophila where the control of cell number and cell fate is critical. Notch is well studied in the wing disc where it interacts with multiple other pathways such as cell death (DIAP1), wnt (wingless) and Hippo signaling (Djiane et al., 2014). Notch signalling takes on a contrasting role in lineage decisions during formation of macrochaete or bristles which cover the head and body of the fly, including the wing, where it first selects a sensory organ precursor via lateral inhibition and then determines one of the SOP daughter cells to give rise to the supporting cell lineage of the shaft and socket of the bristle, whilst the other that receives an inhibitor of Notch, Numb, gives rise to the neural lineage, the neuron and the glial cell (Furman and Bukharina, 2008). Thus the downstream effects of Notch vary throughout development with broad regulation of proliferation at some stages and precise selection of cell fates in others.

As a critical regulator of growth, failures in the normal function of Notch ultimately leads to disease. In mammals, Notch directly regulates a target called Myc through induction of epigenetic changes at the Myc enhancer, namely an increase in the activation mark H3K27 acetylation(Yashiro-Ohtani et al., 2014). Over activation Myc caused bv abberant Notch signalling in T-ALL and breast cancer drives tumourigenesis (Efstratiadis et al., 2014). Myc itself is a bHLH transcription factor that regulates genes that promote tumour growth (Fallah et al., 2017). So prominent is the role of Notch and its target Myc in some cancers that mutations in NOTCH1 drive 65% of cases (Sanchez-Martin and Ferrando, 2017). Another Notch target critical for T-ALL cancers is Hes1, which has context dependent roles in T-cell development and maturation (Wendorff et al., 2010). While direct regulation of Myc and Hes1 drives some cancers, in other contexts the loss of Notch can result in oncogenesis. For example, the development of squamous cell carcinomas follow the loss of Notch signalling and Notch mediated tumour suppressor activity in various epithelial tissues (Nowell and Radtke, 2017). Thus, Notch has diverse context dependent effects in both normal growth, differentiation and diseases mediated through target genes many of which are transcription factors important for growth and differentiation.

1.1.4 The HES gene family respond to Notch

The direct regulation of the Hairy/Enhancer of Split (HES) gene family by Notch is conserved across the animal kingdom. Activation of the genes is mediate by the activating complex comprising Su(H), NICD and Mam at a Su(H) motif, whilst some genes in the family also have enhancers that integrate input from tissue specific proneural transcription factors (Bray, 2006). HES genes encode transcriptional repressors that share conserved features such as a basic DNA binding domain, a helix-loop-helix dimerization domain and Orange domain which is also important for protein-protein interactions (Sun et al., 2007). Whilst some HES proteins such as *Drosophila* Hey possess Y domains important in co-repressor recruitment, E(spl) and Dpn do not (Kobayashi and Kageyama, 2014). Vertebrate genes *Hes1*, *Hes5* and *Hes7* are activated by Notch through Su(H) binding sites in their enhancers (Kobayashi and Kageyama, 2014).

The *Enhancer of Split* (*E(Spl)*) complex is a relatively short stretch of DNA in the *Drosophila* genome with a high density of Notch responsive genes. There are 13 genes within the 60kb *E(spl)* complex *E(spl)-C* , 7 basic-Helix-Loop-Helix-Orange (bHLH-O) type transcriptional repressors, 4 bearded family members (BFM) and a transcriptional co-repressor, all of which are direct targets of Notch (Delidakis et al., 2014). The BFM group genes within *E(spl)-C* harbour Su(H) binding sites in their enhancer regions, respond to Notch signalling, and modify Notch mediated lateral inhibition when misexpressed (E C Lai, 2000). The genes of the bHLH-O group are better characterised and appear to be critical in relaying the Notch signal. Su(H) is bound in the absence of the Notch signal at some strongly responding regions such as the enhancers of *E(spl)-m3* and *E(spl)-m*β and the fastest responding are activated within 5-10 minutes (Housden et al., 2013). Mutations in the coding regions of these genes give rise to neurogenic phenotypes mimicking mutations in Notch itself (Delidakis et al., 2014). The bHLH-O genes in *E(spl)-C* are short intronless genes, whilst other related bHLH-O Notch targets such as *dpn* and *Hey* reside on different chromosomes, possess multiple exons and like the *E(spl)-C* based bHLH-O factors play critical and distinct roles in neural development.

1.1.5 Notch targets and roles of Notch in neuroblast lineages.

Neuroblasts are the neural stem cells of *Drosophila*. Embryonic neuroblasts originate from the neuroepithelium where a subset of equipotent cells are selected to express proneural genes via Notch mediated lateral inhibition (Egger et al., 2008). These cells divide asymmetrically to produce a ganglion mother cell (GMC) and another neuroblast. GMCs divide once and differentiate into the neurons of the larva. The majority of the embryonic neuroblasts undergo programmed cell death. The remainder undergo a period of quiescence before reactivating and populating the central nervous system of the larva with neural cell types including two types of larval neuroblasts (Homem and Knoblich, 2012). Type I neuroblasts are found throughout the brain lobes and the ventral nerve cord and these divide once to produce a single GMC and ultimately two neurons. Type II neuroblasts reside in the central brain and these divide to produce intermediate neural progenitors which then give rise to several GMCs, and ultimately neurons, but in greater numbers than Type I neuroblasts (see Figure 1.3 and also Doe, 2008; Homem and Knoblich, 2012)

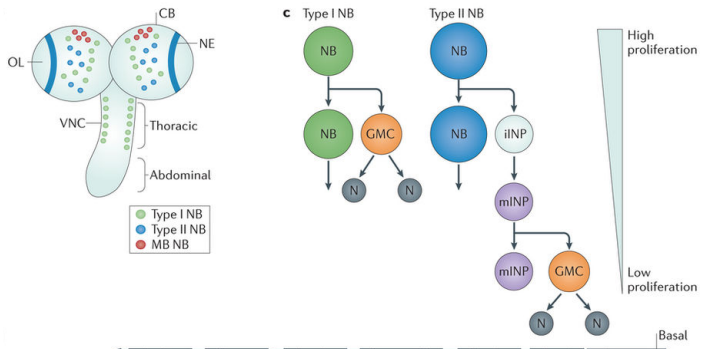


Figure 1.3: Neuroblast lineages of the larval central nervous system (Homem et al., 2015).

The schematic on the left shows the brain lobes above, divided into the optic lobe (OL), the central brain (CB), the neuroepithelium (NE) and the ventral nerve cord (VNC) below. Type I neuroblasts (green), type II neuroblasts (blue) and mushroom body neuroblasts (red). On the right, the lineages are depicted, type I neuroblasts generate one GMC and self-renew, whilst type II neuroblasts generate immature and later mature intermediate neural progenitors (iINPs and mINPs respectively) which produce GMCs and self-renew. GMCs produce two terminally differentiating neurons (N).

Both types of neuroblasts require active Notch signalling in order to maintain normal levels of self-renewal (Homem et al., 2015).Notch signalling is also required for neuroblasts to regrow in between repeated rounds of asymmetric cell division (Homem and Knoblich, 2012; San-Juán and Baonza, 2011). The Notch signal is received via ligands expressed on the surface of their progeny cells. The Notch targets *dpn* and *E(spl)-mγ* are expressed in neuroblasts via Notch regulated enhancers (Zacharioudaki et al., 2012), though there may be some Notch independent regulation of *dpn* (Zhu et al., 2012). All neuroblasts of the larval *Drosophila* brain express *dpn* and ectopic expression causes the progeny of some neuroblasts, intermediate neural progenitors, to revert to a neuroblast like fate and over-proliferate (San-Juán and Baonza, 2011). In addition to neuroblasts, *dpn* is also expressed in the *Drosophila* wing and eye imaginal discs (Babaoğlan et al., 2013). Notch and *dpn* have also been implicated later in development in the cell fate selection in the *Drosophila* eye (Mavromatakis and Tomlinson, 2016) a situation where both the integration with RTK signalling described earlier and the transcriptional repressor properties of *dpn* come into play. When mutated alongside *E(spl)*, *dpn* is found to be required for normal proliferation of neuroblasts (Zacharioudaki et al., 2012). Thus as a Notch target, *dpn* is clearly a critical regulator of cell division in neural stem cells and partly responsible for ensuring the production of the correct number and types of cells during neurogenesis.

Hey also encodes a bHLH-O protein but has distinct roles and expression compared to *E(spl)* and *dpn*. *Hey* is the sole orthologue of a family of vertebrate genes (known as Hrt, Herp, Hesr, Chf or Gridlock) that differ from the *E(spl)* relatives, the Hes genes, due to a characteristic tyrosine (Y) in the C-terminus. Little is known about *Hey* or its relatives, though *Hey* knockout mice have minor defects in neural differentiation and abnormal cardiovascular development. In *Drosophila Hey* is known to be a Notch target but has also been found to be expressed independently in the mushroom body, the brain region that receives inputs from olfactory neurons.

In addition to the Notch-dependent maintenance of stem-cell fate, Notch plays a role in lineage decisions in neuroblast progeny. The two neurons born from a GMC typically have distinct fates determined by which cell receives Notch as they are born (Skeath and

Doe, 1998). In embryonic lineages that have been studied the Notch receiving 'A' type sibling cell goes on to express *Hey* , whilst the sibling with a distinct fate, the 'B' type cell does not (Monastirioti et al., 2010; Ulvklo et al., 2012). Expression profiling in separated cell populations from the CNS has confirmed that *Hey* is highly expressed in neurons and expressed at much lower level in neuroblasts, a stark contrast with *dpn* for which the reverse is true (Knoblich et al., 2012).

In summary, the Notch pathway has many diverse roles in metazoan development and in order to regulate diverse processes it induces different gene expression programmes depending on the context. This is illustrated by the Notch targets described thus far, whilst *E(spl)* and *dpn* are required for normal proliferation of neuroblasts, *Hey* is activated in a Notch dependent manner in terminally differentiated neurons. How these context dependent effects on gene expression are brought about is unclear, and so new perspectives and new tools are required to address possible explanations.

1.1.6 Notch and the chromatin landscape

Epigenetic factors such as the compaction of chromatin, modifications of histones and DNA accessibility can all affect the ability of an individual gene to be transcriptionally activated. Studies in *Drosophila* have categorised chromatin into stereotypical states based on whether they contain active or inactive genes, and whether they are assocatied with proteins that promote or repress gene activity (Filion et al., 2010). Similar approaches have been deployed to determine the chromatin states favoured by Su(H) and epigenetic changes that coincide with Notch target gene activation (Skalska et al., 2015) and Su(H) itself has been found to occupy different binding sites in different cell types(Terriente-Felix et al., 2013). Thus molecular characterisation of chromatin provides a set of circumstances required for Notch gene activation but there have been few efforts to tie this in with the growing field of research regarding nuclear organisation. Moreover, many of the classical context dependent responses to Notch, such as those described in the neural lineages of *Drosophila* occur in cells that are not as easily accessible with genome-wide chromatin profiling as the cell lines, meaning

that new tools will be required to address complex questions of gene regulation in these cell types.

Given Notch has a relatively simple transduction procedure, but many diverse outcomes, it seems feasible that the interpretation of the signal by the genome in the nucleus is a key factor in the varied response to the pathway. How this comes about is not yet fully understood, but the study of nuclear organisation has greatly transformed our understanding of gene regulation in recent years and could hold some of the answers.

1.2 Nuclear organisation and gene expression

1.2.1 Early observations and theories of nuclear organisation

Since early observations of the complex organisation within the nuclear interior, the roles of nuclear landmarks and 3D organisation of chromosomes have become increasingly prominent in our understanding of gene regulation. Electron micrographs of nuclei, such as the one in Figure 1.4, reveal one higher order of organisation in the distinction between lightly staining euchromatin and darkly staining heterochromatin

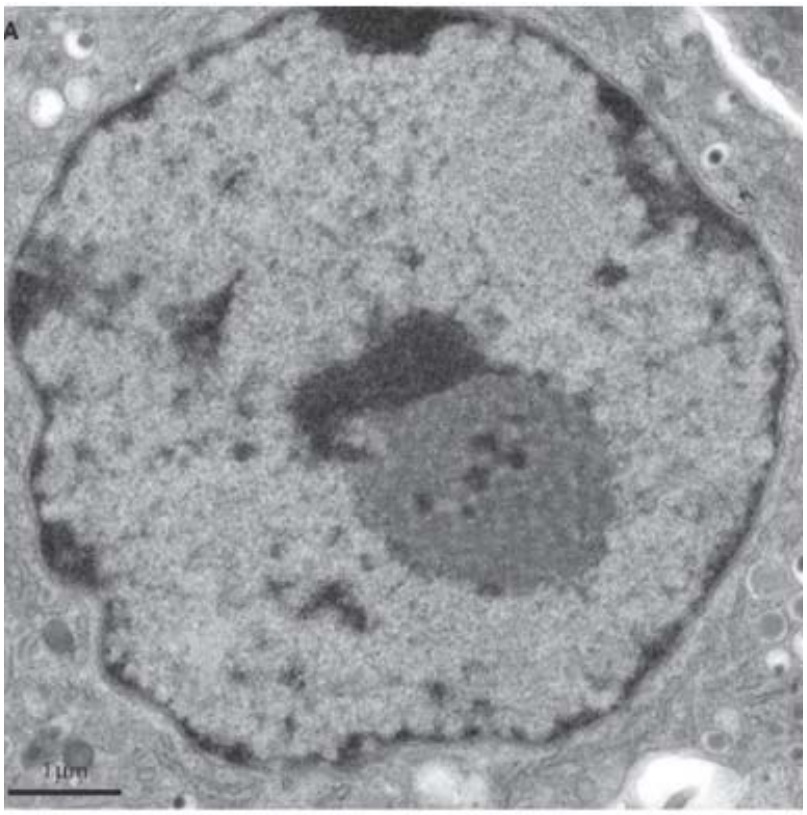


Figure 1.4: Electron micrograph of a mammalian liver nucleus (Akhtar and Gasser, 2007)

Darkly staining heterochromatin is visible in the nuclear periphery at associated with the inner surface of the nuclear envelope, or lamina. Peri-nucleolar heterochromatin is also visible in the centre of the nucleus. The lightly staining regions in the nucleoplasm and interrupting the peripheral heterochromatin are euchromatin, the latter indicating the presence of a nuclear pore.

In the mid 1980s it was theorized that the genome adopts a different 3D shape in differently differentiated cells and that the information required to direct this is contained within the genome itself. The nuclear landmarks (nuclear pore and lamina for example) were suggested to be responsible for interpreting this information and directing 3D organisation of the nucleus (Blobel, 1985). More recently the nucleus has been thought of as being comprised of numerous subnuclear compartments with distinct

roles. Nuclear pore proteins have been found to be important for gene activation in several species (Capelson et al., 2010; Mendjan et al., 2006; Raices and D'Angelo, 2012) and gene nuclear pore interactions have been found to be important in gene activation (Ahmed et al., 2010; Pascual-Garcia et al., 2017; Rohner et al., 2013) meaning the nuclear periphery is in some instances a gene activating suborganelle.

The location of nuclear pores in the nuclear envelope means that they are surrounded by another nuclear subcompartment which is repressive, the nuclear lamina. In fact lamina associated genes, determined using a DNA methylase (Dam) fused to a Lamin protein were found to be generally silent and late replicating in *Drosophila* cell lines (Pickersgill et al., 2006). The repressive role of the lamina is conserved in mammals (van Steensel and Belmont, 2017) and lamina associated domains (LADs) are 10kb – 10Mb long in mice and humans and enriched in repressive heterochromatin marks such as H3K9me2 and H3K9me3. One study has argued that LADs essentially extend beyond the nuclear envelope as it found that the peri-nucleolar heterochromatin and peri-centromeric heterochromatin act as one functional repressive environment for housing low expressed genes and gene poor regions of the genome (Ragoczy et al., 2014). However, subtle differences are emerging between the peripheral heterochromatin of the lamina and the peri-nucleolar heterochromatin such as the lack of LaminB1 and B2 in peri-nucleolar heterochromatin, but LaminA associated with both compartments (Kind and van Steensel, 2014).

The term ‘transcription factories’ resulted from the observation that transcription does not appear to take place homogenously throughout the nucleus but in concentrated foci, and examples of this phenomenon have found that in some cases co-regulated genes from distinct chromosomes can be found colocalised and sharing transcriptional machinery such as polymerase and common transcription factors (Rieder et al., 2012; Schoenfelder et al., 2009). Genome-wide approaches to measuring genome organisation are now being employed to map contacts between different loci in three dimensions and unearth the factors that are responsible for coordinating the packaging of an entire genome into a space only several microns across. These methods have revealed that there are some common principles in solving this problem across the tree of life. For example yeast, invertebrates and mammals all display topologically

associating domains (TADs). In spite of these similarities the factors and motifs that marshal the genome into these conformations are less well conserved (Rowley and Corces, 2016).

1.2.2 Chromatin conformation capture

The development of chromatin conformation capture technologies has revealed genome-wide, population-level maps of genome organisation, the conservation of TADs across many species, and the links between nuclear architecture and gene expression. 3C uses formaldehyde fixation of cells followed by restriction enzyme digestion and ligation to create chimeric DNA fragments representing 3D contacts between regions of genomic DNA that lie in close 3D proximity, enabling the frequency of contacts between loci in cis or in trans from one another to be determined. These details are inferred through bioinformatics following the sequencing of the pool of chimeric molecules, with statistical approaches to correct for the high frequency of contacts between nearby loci. The application of high-throughput next generation sequencing to this approach, known as 'Hi-C', has enabled the production of genome wide maps of contact frequencies between every region of the genome against every other region (de Wit and de Laat, 2012). One key result from many Hi-C studies is that the genome is folded into TADs that appear as regions of relatively dense contacts, typically biased for nearby regions and bounded by insulator proteins (Dekker and Mirny, 2016; Gibcus and Dekker, 2013). TADs have been found to correlate with gene expression with more dense TADs representing compacted environments with lower average expression levels, whilst TADs with a lower contact frequency are inferred to represent more open chromatin and are generally found to be more transcriptionally active(Lieberman-Aiden et al., 2009; Schwartz and Cavalli, 2017). The chromatin conformation capture approaches have led to the domain model of genome organisation that proposes that the genome is subdivided into functional units averaging 100kb in length (in *Drosophila*, 1Mb in mice) that overlap considerably with the bands of polytene chromosomes and the key epigenetic signatures described as chromatin states (Sexton et al., 2012; White, 2012). Thus, the organisation in the nucleus underpins gene regulation and is therefore critical to genome function.

1.2.3 Nuclear architecture of larval salivary gland nuclei

The nuclei of the Drosophila larval Salivary gland contain giant chromosomes generated through a process of endoreduplication, or replication without separation, resulting in up to 1024 closely aligned chromosome copies (Mortin and Sedat, 1982). The polytene chromsomes have long been a popular model for studying genome organisation. Physical analysis of chromosome shape has found that polytene chromosomes have a strongly chiral organisation, predominantly twisted in a right-handed coil configuration and certain heterochromatic domains are consistently associated with the nuclear envelope (Hochstrasser et al., 1986). The arms of the autosomes were found to align alongside one another, but the remainder of the organisation was found to vary widely. Detailed studies of the ultrastructure of the chromosomes found that they are circular in cross section whether one or two homologous chromosomes is present and that each circular cross-section is comprised of approximately 36 smaller bundles of chromatids (Urata et al., 1995). The chromosome banding pattern has been found to overlap considerably with the TADs in a Hi-C based study (Eagen et al., 2015) and substantial similarities have been observed between the banding pattern and the equivalent chromosome conformations in non-polytene cells (Eagen et al., 2015; Vatolina et al., 2011). Recently, modelling approaches have been used to determine the relationship between the number of chromosome to nuclear envelope contacts and inter and intra-chromosomal contacts revealing that more contact points with the nuclear envelope results in increased intra-chromosome and intra-arm contacts whilst reducing inter-chromosome and inter-arm contacts. Thus, the contact points between the polytene chromosomes and the nuclear envelope are instrumental in regulating the genome organisation of salivary gland nuclei (Kinney et al., 2015).

1.2.4 Insights into genome organisation from DNA-FISH studies

The localisation of specific chromosomes and individual genes is typically observed via DNA fluorescent in situ hybridisation (DNA-FISH) or locus-tagging approaches. DNA-FISH, a method for labelling specific genomic sequences in cytological studies was developed in the late 1960s (Gall, 2016) initially focusing on repetitive sequences and later duplicated polytene chromosomes (Langer-Safer et al., 1982) due to the amplification of the target site. The availability of commercial probes and fluorescent dyes now enables precise super-resolution imaging of specific loci and also accurate quantification of local expression when applied to RNA species (Boettiger et al., 2016). The combination of chromosome conformation capture methods and DNA-FISH has been used in *Drosophila* embryos to show that the repressive polycomb family proteins drive a 3D organisation referred to as 'gene-kissing'. When a sequence element required for normal 3D contacts, Fab-7, was mutated, nearby genes were derepressed, supporting the notion that 3D organisation of the genome is important for gene regulation (Bantignies et al., 2011). Using a similar combination of approaches, DNA-FISH and 5C, tissue and developmental stage specific contacts between the *Shh* gene and its enhancer ZRS have been mapped, illustrating that the organisation of the genome can dynamically readjust through animal development to drive highly specific gene regulatory events (Williamson et al., 2016). The correlation between gene position and expression has also been directly tested in the *Drosophila* embryo, where a developmental transcription factor, hunchback was found to move to the periphery and increase colocalisation with the lamina during the same developmental stages in which it undergoes a downregulation, (Kohwi et al., 2013). A key limitation of FISH is that the tissue is fixed prior to application of the probe and high temperatures in the hybridisation procedures mean that it is not always straightforward to combine with immunostaining.

1.3 Locus tagging

1.3.1 Locus tagging systems utilizing LacO/LacI-GFP

Fortunately, other approaches exist, such as locus tagging with genetically encodable labels. Locus tagging was developed by yeast biologists in the 1990s by inserting repeats of the Lac operator sequence into the yeast genome allowing them to visualise chromosome dynamics in living cells when a LacI-GFP fusion protein accumulated on the array of repeats (Belmont and Straight, 1998; Straight et al., 1996).The technique has been adapted by several other model organisms including nematodes and plant species, and a variant based on the analogous TetO/TetR system has also been developed (Jovtchev et al., 2011; Matzke, 2005; Tanenbaum et al., 2014). In the yeast model system where homologous recombination is highly efficient LacI-GFP/LacO has been used to provide evidence for nuclear pore-mediated gene-gating of certain metabolic genes following a change in carbon source and DNA-Zip codes that direct these reorganisations of the genome (Ahmed et al., 2010; Brickner et al., 2012). An analogous phenomenon has been observed in the nematode *C.elegans* where a heatshock promoter was fused to a LacO repeat and inserted into the nuclei of embryos. Following a heatshock the subnuclear localisation was found to be strongly biased towards the periphery and co-localised with a nuclear pore marker suggesting that nuclear pore directed genome reorganisations could be a common feature of rapidly responding genes (Rohner et al., 2013). In order to gain a more comprehensive understanding of gene dynamics, locus tagging has been combined with high-throughput imaging and automated analysis to build a picture of the movements of genes over time(Berger et al., 2008; Wang et al., 2016a ; Wang et al., 2016b). This is a powerful approach because of the various sub compartments described above and the dynamic reorganisations that occur in the nuclei of the majority of species and cell types studied thus far. These studies have found that genes occupy specific territories and that nuclear landmarks, such as the nucleolus are important factors in determining genome organisation and dynamics.

1.3.2 The ParB/Int locus tagging system

Recently a novel locus tagging system was presented, again pioneered by researchers studying DNA dynamics in yeast, but with a number of features making it preferable to the original LacI-GFP/LacO system. The LacI-GFP/LacO system was initially chosen due to the strong affinity between the protein and DNA binding motif (Straight et al., 1996), however this means that this locus tagging system may create an artificial barrier to processes that require the ability to 'read through' the tagged region. The novel system is based on the ParABS chromosome partitioning proteins from *Burkholderia cenocepacia* and their recognition motif 'parS' sequences (Dubarry et al., 2006). The key advantage lies in the fact that fluorescent foci are produced by protein oligomerization, rather than by multiple copies of the DNA-binding motif. Thus the majority of the fluorescent protein is localised in a non-sequence specific way (ParB spreads along the chromosome after initial recognition) and further characterisation confirmed that this means the locus tag is readily displaced and dominated by transcriptional processes and also that normal nucleosome formation was detected in the region of the insertion (Saad et al., 2014). Finally, the ParB-Int locus tagging system is based on two distinct protein-DNA interactions that do not cross-react and thus it represents a less intrusive, dual colour alternative to the LacI-GFP/LacO locus tagging system.

1.4 Aims

This thesis aimed to develop a CRISPR/Cas9 based approach to retarget the ParB-Int locus tagging system. A key goal was to characterise the ParB-Int system in multiple cell types with existing control Int insertions and novel insertions at intergenic regions between *E(spl)* loci, adjacent to *dpn,* and *Hey*. Given nuclei vary in size and shape and gene position can only be determined relative to the nucleus in question another goal was to develop a systematic quantification approach that could detect nuclear dimensions and gene position as highlighted by the locus tag. Ultimately, the goal of this project was to determine the way in which Notch responsive genes are organised in

3D space and to ask whether any of the tagged loci are non-randomly positioned and reside in specific gene territories.

2 Methods

2.1 Molecular Biology

2.1.1 PCR

Polymerase chain reaction was performed using NEB taq (genotyping, colony PCR) or Q5 polymerase (cloning genomic fragments, Gibson Assembly inserts) as per the manufacturer's instructions.

2.1.2 gRNA plasmid construction

Guide RNA sequences for the desired loci were selected using the gRNA design tool at http://www.flyrnai.org/crispr/ and filtered for gRNA efficiency score of 6 and above. Single guide RNA plasmids were cloned using the protocol from Fillip Port available at http://www.crisprflydesign.org/ and referred to in (Port et al., 2014). pCFD3 was cut with BbsI restriction enzyme (NEB) and cleaned with a PCR purification column (QIAGEN). A pair of complimentary gRNA oligos were designed as per the protocol containing 20 bases of complimentary sequence and 4 bases at the 5' end of each oligonucleotide to complement the cut ends of the vector. The oligo pair was annealed in a thermocycler with the ramp down specified in the protocol. The resulting inserts were diluted and added to a ligation reaction (T4 ligase, Thermofisher) along with 50ng of the cut vector backbone. Ligation was performed at 18°C for one to two hours or overnight.

2.1.3 Homology directed repair template construction

Homology templates were based on the pHD-DsRed.attP plasmid from the O'Connor-Giles lab (http://flycrispr.molbio.wisc.edu/protocols). For the *ebony* HDR template, the 5' homology arm was amplified from Kc167 cell line genomic DNA (extracted with a QIAGEN DNEasy kit as described in 2.1.7). See appendix 1 for oligonucleotide list. The 5' homology arm inserted using EcoRI and NotI restriction enzymes. The 3' arm of the *ebony* HDR template was amplified from genomic DNA prepared from the fly strain pCFD2 as described in 2.1.7 and was inserted using SapI restriction enzyme (NEB). All other HDR template components derived from genomic sequence (homology arms for E(spl), dpn and Hey HDR templates) were amplified from pCFD2 strain genomic DNA as above. Primers were designed to amplify homology arms (~1000bp in length) 5' and 3' to the first 6bp of the gRNA target site. Enzyme cut sites plus 4-6 bases were added to the 5' end of the oligonucleotides to enable digestion of the PCR products. Each homology arm was added sequentially, with the plasmid being recovered by colony PCR and verified by Sanger sequencing. The remaining HDR plasmids (*E(spl)-mδ, E(spl)-m8*, *dpn*, *Hey*) were constructed using Gibson Assembly. In all cases primers were designed to clone ~1000bp of genomic DNA For Gibson Assembly, the primers were designed to have at least 40bp of overlap between the adjoining fragments and at least 20bp from each of the two fragments to be joined. All HDR template plasmids retained the attP site and DsRed-Express expression cassette from the original pHD-DsRed.attP plasmid and the AmpR resistance cassette.

2.1.4 Oligonucleotides

Please see appendix 1 for a list of oligonucleotides used in gRNA construction, homology-directed repair template construction and genotyping.

2.1.5 Transformation of *E.coli*

Ligated products and completed Gibson Assembly reactions were transformed into chemically competent *E.coli* (Mach1) via 30 minute incubation with the ligation on ice followed by a 30 second heatshock at 42°C. The cells were recovered with SOC prewarmed to 37°C and plated on LB-Agar plates (containing Ampicillin at a concentration of 100µg/ml) that were prewarmed to 37°C.

2.1.6 DNA electrophoresis

DNA fragments were sized using Gene Ruler Plus 1Kb (Thermofisher) run together on 0.5% or 1% agarose gels. Agarose was mixed with TAE buffer and melted in a microwave. Gels were run in TAE buffer using a Biorad gel tank using 120V for 10-20 minutes. Following electrophoresis, fragments were visualised by submerging the gel in an ethidium bromide and double distilled water solution (final concentration 0.5µg/ml) and imaged on a gel doc. In the case of gel extractions (see below), bands were excised using a scalpel on a UV transilluminator.

2.1.7 DNA extraction

PCR products and restriction digested DNA fragments were purified with PCR purification kits (QIAGEN) or gel extraction kits (QIAGEN) as per the manufacturer's instructions. For genomic DNA the DNeasy Kit was used following the manufacturer's recommendations for insect tissue preparation.

2.2 Immunohistochemistry

2.2.1 Sample preparation

Wandering third instar larvae were dissected in phosphate buffered saline. The carcass was inverted using forceps and immediately fixed in 4% PFA for 10 minutes with shaking, washed for 5 minutes in PBT (0.3% Triton), permeabilised with PBT (1% Triton) for 20 minutes and blocked for 1 hour with BBT (0.3% Triton & Bovine Serum Albumin). Primary antibodies diluted in BBT were incubated with the carcasses overnight at 4°C. Primary antibodies were washed off with three washes of 15 minutes in BBT. Secondary antibodies diluted in BBT were incubated with the carcasses for three to four hours at room temperature with shaking. Secondary antibodies were washed off with three washes of 15 minutes using PBT (0.3% Triton) and the carcasses were finally incubated with Hoechst 33342 (Sigma) at a final concentration of 20mM for 10 minutes before a final wash in PBT (0.3%Triton).

2.2.2 Primary Antibodies

anti-Lamin (Developmental Studies Hybridoma Bank) [1 in 50]

anti-*dpn* (gift from C. Delidakis, unpublished [1 in 100]

anti-Hey (described in (Monastirioti et al., 2010), gift from C. Delidakis) [1 in 100]

2.2.3 Secondary Antibodies

(all 1 in 200)

anti-Mouse FITC (Jackson Immunoresearch/Stratech)

anti-Mouse Cy3 (Jackson Immunoresearch/Stratech)

anti-Guinea Pig Cy5 (Jackson Immunoresearch/Stratech)

anti-Rabbit Cy5 (Jackson Immunoresearch/Stratech)

anti-Guinea Pig Alexa 555 (Molecular Probes/Life Technologies)

2.2.4 DNA stain

25mg of Hoechst 33342 (Sigma) was resuspended in 500µl ddH_2O and kept in the dark at 4°C. A 1 in 250µl dilution was made into PBS with 0.3% Triton to give a final concentration of 200µg/ml. Samples were incubated with this at room temperature for 5-10 minutes after the final wash step of antibody staining.

2.2.5 Mounting

Following staining carcasses were dissected in PBT (0.3% Triton) to free either the salivary gland or central nervous system from debris. Slides were prepared with a tape spacer (Sellotape, double sided) for consistent spacing and to prevent flattening of the tissue., 2-3 3/8" holes in the tape were made with a hole punch to enable 2 or more samples to be imaged on the same coverslip. Tissue samples were transferred with a pipette and tissue paper was used to blot away excess liquid without touching the samples. 12 µl of Vectashield was added to the well immediately and a 60mm x 24mm coverslip was placed on top and gently sealed with light pressure from forceps. Mounted samples were stored at 4°C and typically imaged within 1-2 days of preparation.

2.3 Imaging

All confocal imaging was performed on a Leica SP8 using a 63X oil immersion objective with a numerical aperature of 1.4 (HC PL APO CS2 63x/1.40 OIL) unless specified otherwise. Excitation and emission wavelengths were the default options selected based on the closest match in the LAS X wizard, fluorophores used were:

Fluorophore	**LAS X Wizard**	**Excitation (nm)**	**Emission Band Pass (nm)**
Hoechst 33342	Hoechst 33258	405	420-500
mCherry	mCherry	561	607-683
GFP	EGFP	488	500-550
Cy3	Cy3	545	560-580
Cy5	Cy5	594	650-710
Alexa555	Alexa555	555	570-590

Salivary gland nuclei were imaged with 2µm step spacing in Z and a 2X digital zoom applied for all acquisition. Neuroblast nuclei were imaged with a 0.4 µm spacing in the Z and a 6X digital zoom applied for all acquisition.

2.4 Fly husbandry

2.4.1 Fly food

Flies were reared on a mix of cornmeal, yeast, agar and glucose, with a preservative and a blue food dye as per the following recipe:

Glucose : 76g/l
Cornmeal flour : 69g/l
Yeast : 15 g/l
Agar : 4.5g/l
10% p- hydroxy-benzoic acid Methyl Ester in Ethanol : 25ml/l
(preservative agent known as Tegosept/Nipagin with
1% Blue food colour dye from Ingram Brother)

2.4.2 Temperature control

Stocks were typically maintained at room temperature, 18°C or 25°C. Crosses using inscutable-Gal4 ~ gal80^{ts} were shifted to 29°C 24-48hours after egg laying to trigger inhibition of the gal80^{ts} and permit the activity of Gal4 to activate the UAS element driving expression of any UAS fused transgenes, typically ParB1-GFP or ParB2-GFP.

2.5 Locus tag quantification with FIJI and R

2.5.1 Image processing tools used

FIJI & ImageJ:

Version: 2.0.0-rc-59/1.51n

Build: fab6e1a004

Open source image processing software

Copyright 2010-2018

http://imagej.net/Contributors

3D Suite plugins (including 3D object counter, 3D viewer and 3D ROI manager) by Thomas Boudier:

URL: http://imagejdocu.tudor.lu/doku.php?id=plugin:stacks:3d_ij_suite:start

Installation Bundle: mcib3d-suite3.9

2.5.2 R programming

R project:

R Core Team (2016). R: A language and environment for statistical computing. R Foundation

for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Source: https://cran.r-project.org/

Packages used:

Rgl (3D plotting)

reshape2 (reformatting data)

vioplot (violin plots)

nortest (normality test)

User defined functions (to be run in R):

Distance-3D-function:

```
#Function that takes two vectors of x,y,z co-ordinates for two points in a 3D
#array. It subtracts all the co-ordinates and takes the modulus, giving the
#edge distances as a vector of three values - two results from this can be put
#into pythag3D to calculate the longest diagonal between the two points.

distance3D = function(b,c){
    x = abs(b[1] - c[1])
    y = abs(b[2] - c[2])
    z = abs(b[3] - c[3])
    d = c(x,y,z)
    names(d) = c('x','y','z')
    return(as.numeric(d))
}
```

Pythag 3D function:

```
#A function that does pythagorus in 3D given the three vertices of the cuboid
#for which the longest internal diagonal is required.
pythag3D = function(x,y,z){
    w = sqrt(x^2+y^2+z^2)
    return(w)
}
```

2.5.3 Genome and plasmid visualisation

In order to produce the diagrams of the genome in Figures 3.2A, 3.4A and 4.1, a genome browser named IgV developed and maintained by the Broad Institute was used. *Drosophila melanogaster* genome assembly version dm3 was selected (whilst dm6 is available in IgV, other datasets used in this thesis were published whilst dm3 was current, so dm3 was used for consistency). Tools>Find Motif was selected and where

necessary gRNA target sequences were pasted into the menu and the results used to as the basis for a diagram of the gRNA target region in genomic context.

In order to visualise plasmid sequences and perform alignments of sequence traces to plasmids and genomic sequences a free online molecular biology tool package named Benchling was used (www.benchling.com). Reference sequences were loaded using the built in menus and in the case of performing alignments, .ab1 files provided by Source Bioscience from Sanger Sequencing services were uploaded into the Benchling alignment tool. Sequence traces in Figure 3.5 and Appendix 2, Figure 8.1D and 8.1E were generated in Benchling from .ab1 chromatogram files and annotations were traced over in Adobe Illustrator CS6 to make the text legible.

2.6 qPCR of *Drosophila* salivary gland RNA.

Wandering third instar larvae were dissected in PBS and salivary glands were transferred to an eppendorf tube and kept on ice. Glands from 40-60 individuals were homogenised in 200µl of Trizol with a microfuge tube pestle. The RNA was then extracted and precipitated as per the manufacturer's instructions (Ambion) , RNA yield was quantified on a nanodrop and then the samples were stored at -80°C until reverse transcription. The Moloney Murine Leukemia Virus Reverse Transcriptase kit from Promega was used to produce cDNA from the 2µg of extracted RNA per sample following per the manufacturer's instructions. The cDNA product for each sample was diluted to 700ng/µl with RNase free water and this template was then used in a qPCR reaction. A Roche Lightcycler 480 II was used in combination with the Roche Sybr Green master mix (5µl master mix : 0.3µl forward primer (10µM) : 0.3µl reverse primer (10µM), 3.4µl ddH_2O : 1µl of template cDNA). Primers used are detailed in the oligonucleotides table (Appendix 1) Relative amounts of cDNA were calculated using the ddct method (Jitao David Zhang, Rudolf Biczok and Markus Ruschhaupt (2015). ddCt: The ddCt Algorithm for the Analysis of Quantitative Real-Time PCR (qRT-PCR). R package version 1.30.0.).

2.7 Neuroblast culture method

(For solutions, see Appendix 4)

Larval L3 brains were dissected in the dissecting solution, transferred to collagenase solution, and incubated for 15 minutes. Collagenase was removed and the brains were rinsed three times with culture medium. 40μl of medium/brain solution was transferred into an eppendorf tube and dissociated into single cells by pipetting up and down. Cells were rested for 30 minutes on Poly-D-Lysine coated plates prior to imaging.

3 Targeting a locus tagging system to Notch responsive genes

3.1 Introduction

3.1.1 Locus tagging approaches based on DNA binding motif repeats

In order to determine the subnuclear localisation of Notch responsive genes, tools are required that label the loci of interest and enable quantification of their localisation. DNA-FISH has been used in many studies to determine the location of a certain genomic sequence, locus or whole chromosome (Boyle et al., 2011; Cremer and Cremer, 2006; Pecinka et al., 2004). However, combining this methodology with immunostaining, for example to identify specific cell types or subcellular structures, is not always reliable. Furthermore DNA-FISH requires nuclei to be fixed and the DNA to be denatured so it is not compatible with *in vivo* live imaging. Strategies to specifically detect a genomic region *in vivo* with a fluorescent protein, so-called 'locus tagging' systems enable robust, reproducible visualisation of a specific genetic locus and are thus very valuable for more dynamic analysis of nuclear organisation.

The most widely used approach for locus tagging, the LacI-GFP/LacO system was pioneered in the late 1990s as a way to visualise the position and dynamics of chromosomes in live yeast cells (Belmont and Straight, 1998). The bacterial repressor protein LacI and its cognate binding motif the Lac Operator (LacO) were chosen as they were well characterised and known to have a strong binding affinity and proven functionality in eukaryotic cells. The approach relies on a LacO repeat array bearing 256 copies of the DNA binding motif to recruit hundreds of LacI-GFP molecules to the chromosomal locus of interest. The LacO repeat array must be inserted into the genome at the desired location. *Saccharomyces cerevisiae* was an ideal model for this as it is known to have a high rate of recombination when transformed with exogenous DNA bearing a sequence homologous to the desired insertion point.

The LacI-GFP/LacO chromosome labelling approach has been widely successful (Jovtchev et al., 2011; Rohner et al., 2013; Wang et al., 2013). For example, dynamic reorganisations of the genome in yeast were detected in response to changes in gene expression following changes in nutrient availability (Ahmed et al., 2010). A similar strategy demonstrated that elements of the heatshock promoter in the *C.elegans* genome are necessary and sufficient for the translocation of a plasmid containing the heatshock promoter to the nuclear envelope within 30 minutes of a heatshock to the embryo (Rohner et al., 2013). Chromosome labelling approaches analogous to the LacI-GFP/LacO system have been developed such as one using the Tetracycline repressor and operator sequence (Abruzzi et al., 2006). However, key limitations of all these systems are linked to the use of binding motif multiplicity to amplify the fluorescent signal from the LacI-GFP which requires long and repetitive DNA. Moreover, non-native and sequence specific protein-DNA interactions such as that between LacO and LacI-GFP may perturb the local chromatin environment in unknown ways. In addition, although an analogous system using another bacterial binding motif/DNA binding protein pair, TetO/TetR, has been developed and the two systems have been successfully combined, dual labelling of two loci is still hampered by the above limitations and only rare cases have been successful (Roukos et al., 2014).

3.1.2 The ParB-Int system is an alternative to LacO/LacI and TetO/TetR based locus tagging systems

One potential solution to the limitations of the LacI/LacO and TetR/TetO based approaches comes from DNA binding proteins that are able to amplify the fluorescent signal by means other than recognition of motif repeats. The ParB-Int locus tagging system has this property and was recently adapted from a set of chromosome partitioning proteins from the bacterial species *Burkholderia cenocapacia* (Dubarry et al., 2006). Loci are labelled with a short motif (around 1.5kb) which serves as a recognition site for a ParB protein. The ParB protein is monomeric and freely diffusing in the absence of the recognition site, but on binding to the motif it oligomerises and spreads along the chromosome forming fluorescent foci in the case of ParB proteins fused with GFP or mCherry (Saad et al., 2014). The ParB-Int system has two major advantages over the LacI/LacO and TetR/TetO type systems. Firstly because a short region of DNA can recruit many proteins, the overall size of the fragment providing the recognition sequence is much shorter and less repetitive than the repeat arrays used in LacI/LacO type systems and contains far fewer repeats (typically 3, versus 256 in LacO/LacI-GFP). Thus, the Int motifs are much more tractable for molecular biology and transgenesis. Secondly, a pair of non-cross reacting ParB proteins and Int motifs have been tested together in yeast. This enables two genomic loci to be reliably tagged and visualised in different colours simultaneously. This report will follow the nomenclature used by one of the initial publications that used this system in studying the dynamics of eukaryotic chromosomes which names the proteins ParB1 and ParB2 and their cognate binding motifs Int1 and Int2 respectively.

The ParB-Int system has been successfully used to study the chromosomal dynamics of DNA following a double stranded break in *Saccharomyces cerevisiae* by Kerstin Bystrycky and colleagues (Saad et al., 2014). Importantly, this study found that transcription through a resistance cassette was not significantly altered by the accumulation of the ParB fusion protein at the locus suggesting the new locus tagging system is unlikely to perturb local processes in the chromatin such as transcription which would likely be blocked or perturbed by a protein-DNA interaction with strong

sequence specificity such as that found in the LacO/LacI system. Initial progress using the ParB-Int locus tagging system in *Drosophila* has been made by our collaborators who had generated a set of control transgenic flies with Int1 and Int2 insertions along with strains for expression of the fluorescent ParB1 and ParB2. Their preliminary data indicated they could successfully detect foci from ParB protein recruitment in the epidermis of the *Drosophila* embryo (Francois Payre and Phillipe Valenti, personal communication).

In order to realise the benefits of the ParB-Int system, a method is required to deliver the Int motif to the locus of interest. The emergence of CRISPR-Cas9 genome editing, a significant advance in the simplification of locus specific genome-engineering has made it possible to generate such locus specific insertions of the Int sequences. Therefore a strategy was designed to insert the Int sequences into intergenic regions within a few kilobases of key Notch target genes. The two stage strategy combining CRISPR/Cas9 genome editing and ΦC31 integrase mediated transgenesis is illustrated in Figure 3.1 (overleaf). The novel locus tagged strains will reveal the nuclear location of the tagged gene in a cell type of interest when crossed with a stock expressing the appropriate ParB variant under control of the Gal4/UAS system. As will be shown this strategy enabled this project to achieve single locus resolution of Notch responsive genes and study their subnuclear localisation. The methods developed here will likely be used more widely in future to ask a broader range of biological questions at the single gene level as the ParB system appears to be a powerful and versatile tool kit.

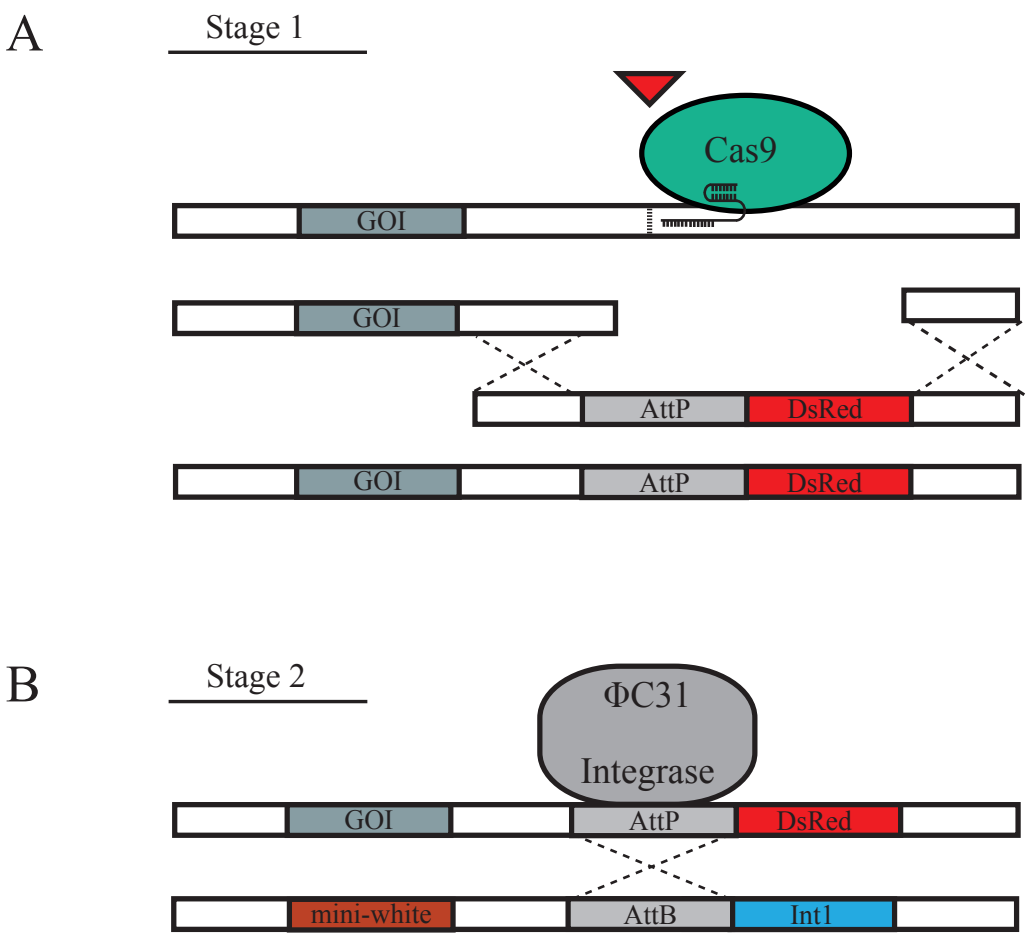


Figure 3.1: A transgenesis strategy for labelling Notch responsive loci with Int1 and Int2 sequences.

In stage 1 a pCFD3 plasmid loaded with a gRNA target sequence is injected into early embryos from the pCFD2 stock which expresses Cas9 under the control of a germline specific promoter (nanos). The gRNA (black hairpin) is recognised by Cas9 and finds the target site in an intergenic genomic sequence next to the gene of interest (GOI). A double strand break is generated in the genome via Cas9 nuclease activity. The homology template aligns with the sequences flanking the break site. The break is repaired via endogenous repair machinery through homology directed repair using the homology template and the attP-loxP-DsRed-loxP cassette is repaired into the break site (A). In stage 2 the stock containing the DsRed cassette and the novel attP site is crossed to a stock expressing ΦC31 integrase under the control of a germline specific promoter. The resulting stock is injected with an attB containing plasmid bearing cargo such as the Int1 motif from the ParB-Int system. Transformants are identified via the mini-white marker which produces a red pigment in the eyes of otherwise *white* $^{-/-}$ flies (B).

3.2 Results

3.2.1 Testing efficacy of CRISPR/Cas9 genome editing

CRISPR/Cas9 genome-editing gives much greater scope to use locus tagging strategies to interrogate the genome. Cas9 is an RNA programmable nuclease resulting in highly specific double strand cleavage of plasmid DNA targeted by a pair of essential short RNAs the crRNA and tracrRNA (Jinek et al., 2012). Combining the two targeting RNAs into a single chimeric sequence now referred to as the guide RNA or gRNA facilitiates rapid and efficient retargeting of the Cas9 nuclease activity and has even enabled the editing of multiple genes in a single transfection (Cong et al., 2013). CRISPR/Cas9 has now been used in a large variety of genome-editing applications in various different model organisms. Plasmid vectors for the expression of gRNAs in *Drosophila* and transgenic fly strains expressing the Cas9 enzyme in the germline have been tested successfully and in the presence of a repair template, homology directed repair works efficiently(Gratz et al., 2014a; Port et al., 2014).

The first aim was to establish and optimise the genome-editing procedure in our lab by reproducing the experiments that targeted the pigment gene *ebony*. The advantage of this approach is that successful editing gives an easily scored visible phenotype in adults. The gRNA from Port and Bullock 2014 maps to the first exon in the *ebony* gene as shown in Figures 3.2A and 3.2B. A red triangle indicates the putative cleavage site as annotated in the original publication (Port et al., 2014). The Cas9 enzyme frequently induces double stranded breaks within the gRNA target sequence which result in indels via the error-prone repair process non-homologous end-joining (Port et al., 2014). The aim of creating indels in the first exon of a protein coding gene is to induce a frameshift mutation resulting in a null phenotype. Mutating a pigment gene such as *ebony* in this way is predicted to produce viable offspring with a clearly visible phenotype, in this case dark body pigmentation.

The pCFD3 expression cassette was used to produce the gRNAs. It contains a U6 promoter driving expression of a gRNA that begins with a 20 nucleotide gRNA target

sequence and is fused to a 76 basepair structural RNA referred to as the gRNA scaffold (see Figure 3.2C) (Cong et al., 2013). After inserting the *ebony* targeting sequence, the pCFD3.ebony plasmid was injected into embryos from the flystock pCFD2 which expresses Cas9 in the germline, under the control of the *nanos* promoter (phenotypic example of the pCFD2 stock is in Figure 3.2D). At first a concentration of 1 µg/µl was used which is a standard injection concentration for plasmids used in ΦC31 integrase mediated transgenesis a site specific genome-engineering method based on bacteriophage recombinases developed prior to CRISPR/Cas9 (Bischof et al., 2007). Surviving flies were crossed individually to the balancer stock w-;;TM3/TM6b (example in Figure 3.2E). Both balancer chromosomes contains a null mutation in *ebony* on each third chromosome, giving the stock a dark body pigmentation which is lost when heterozygous offspring are generated such as when w-;;TM3/TM6b is crossed to stocks with wild type *ebony* alleles. However, in flies where a *de novo* mutation has arisen in the *ebony* allele, the F1 progeny will have an ebony, dark pigmented body colouration phenotype. Novel null alleles for the *ebony* gene were retrieved from the injected individuals based on this phenotype which indicated successful genome-editing. Additionally, the female offspring exhibited an orange eye phenotype intermediate between the red eye phenotype of the pCFD2 stock and the white eye phenotype of the w-;;TM3/TM6b stock which confirmed they were indeed F1 progeny, as shown in Figure 3.2F. The new mutants were thus phenotypically distinct from either of the parental strains shown in Figures 3.2D and 3.2E. This strategy resulted in 5 of the fertile crosses giving rise to progeny with successful knock-out mutations. However, as a proportion of the total number of injected embryos just 0.008% were successfully edited and survived, suggesting improvements could be made to increase the efficiency of the procedure, specifically with regards to increasing the survival.

Due to the poor viability observed in the initial test and the anticipated need to coinject with another plasmid (e.g. for homology directed repair experiments), the experiment was repeated with a lower gRNA concentration. It was reasoned that the high proportion of null alleles in the survivors suggested that the gRNA was very efficient and may be causing a high degree of DNA damage and toxicity. The reduction in gRNA concentration increased viability with only a modest decrease in the proportion

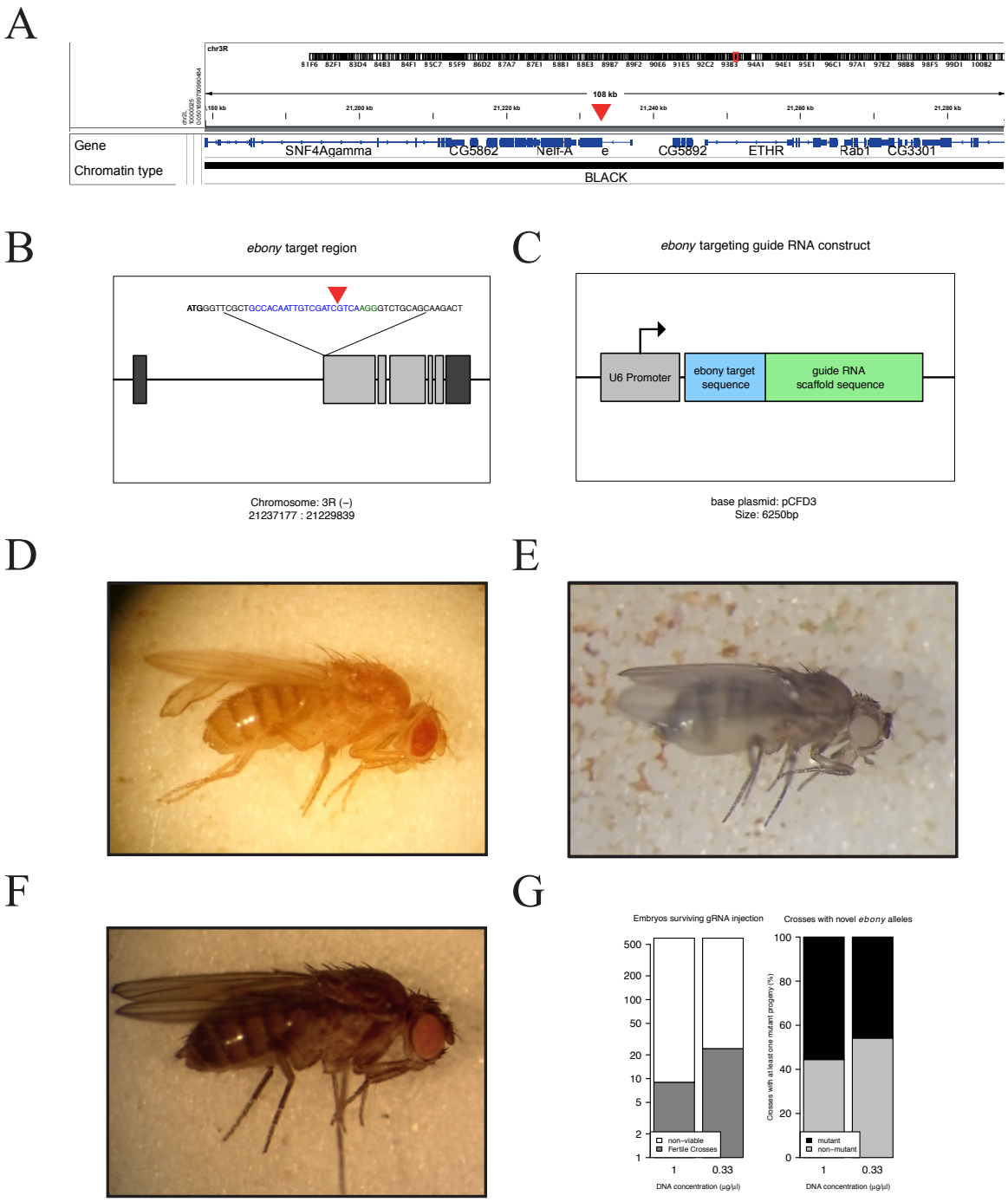


Figure 3.2: Testing CRISPR/Cas9 via a pigment gene knock out.

Genome browser view of the gRNA target site in the first exon of the ebony gene (A). A map of the ebony locus showing the starting sequence of the first exon and the gRNA sequence in blue (adapted from Port and Bullock 2014) (B) . Red arrow marks approximate indel site. Schematic of gRNA expression plasmid based on pCFD3 (C). Example of pCFD2 stock showing eye and body colouration (*white+* and *yellow-*, respectively*)* (D). Example of balancer stock w-;;TM3/TM6B showing eye and body colouration (*white-* and *ebony-)*, respectively) (D). A female fly exhibiting the *ebony* $^{-/-}$ body colouration phenotype and an orange eye colour following (F). (G) Scoring of progeny from the crosses between genome edited males and virgins from the w-;;Tm3/Tm6B balancer stock.

of mutant offspring produced; 1 µg/µl produced 55% mutant offspring with only 9 surviving fertile males, 0.33 µg/µl produced 45% mutant offspring with 24 surviving fertile males (Figure 3.2G).

3.2.2 Homology directed repair at the *ebony* locus.

As the ultimate goal was to insert the Int sequences into the genome by homology directed repair, the next step was to test whether co-injection of the *ebony* gRNA with a homology directed repair template resulted in reproducible transgene insertion at the *ebony* locus. A repair template plasmid was constructed based on the plasmid pHD-DsRed.attP (Gratz et al., 2014b). Homology arms of approximately 1kb were included 5' and 3' to the attP site and DsRed marker cassette (schematic in Figure 3.3A). The final sequence was designed such that no sequences matching the gRNA target remained in the desired insertion. Three rounds of injection were performed to gauge the efficiency and reproducibility of any successful HDR events. The progeny were now scored both for dark body pigmentation resulting from CRISPR/Cas9 induced *ebony* null allele generation and DsRed expression in the eyes of the adult flies indicating the marker plasmid had been integrated into the genome. We opted to only screen for editing in males, because F1 females retained an X-linked DsRed marker from the nos-Cas9 cassettee. Successful insertions were detected in all three biological replicates (Figure 3.3B). Despite the consistent production of successful knock-ins, there was a wide variability in the proportion of knock outs to knock ins in the male germline. Nevertheless the average proportion of fertile crosses with the correct genome-edits varied from 2/44 to 2/25 (4.5% - 8%, mean of three replicates 6.6%).

The insertion location of the HDR template was confirmed by PCR for an amplicon specific to successful fusions between genomic sequence and the inserted transgene. This approach is designed to distinguish between flies with wild type *ebony* alleles and those with the DsRed transgene inserted at the *ebony* locus as well as ruling out integrations at unintended locations. Schematics in Figure 3.3D show the primer annealing sites for reactions (a) a locus distant from the target, *E(spl)-m3* (positive amplification control), (b) a short amplicon to be amplified from the wild type *ebony*

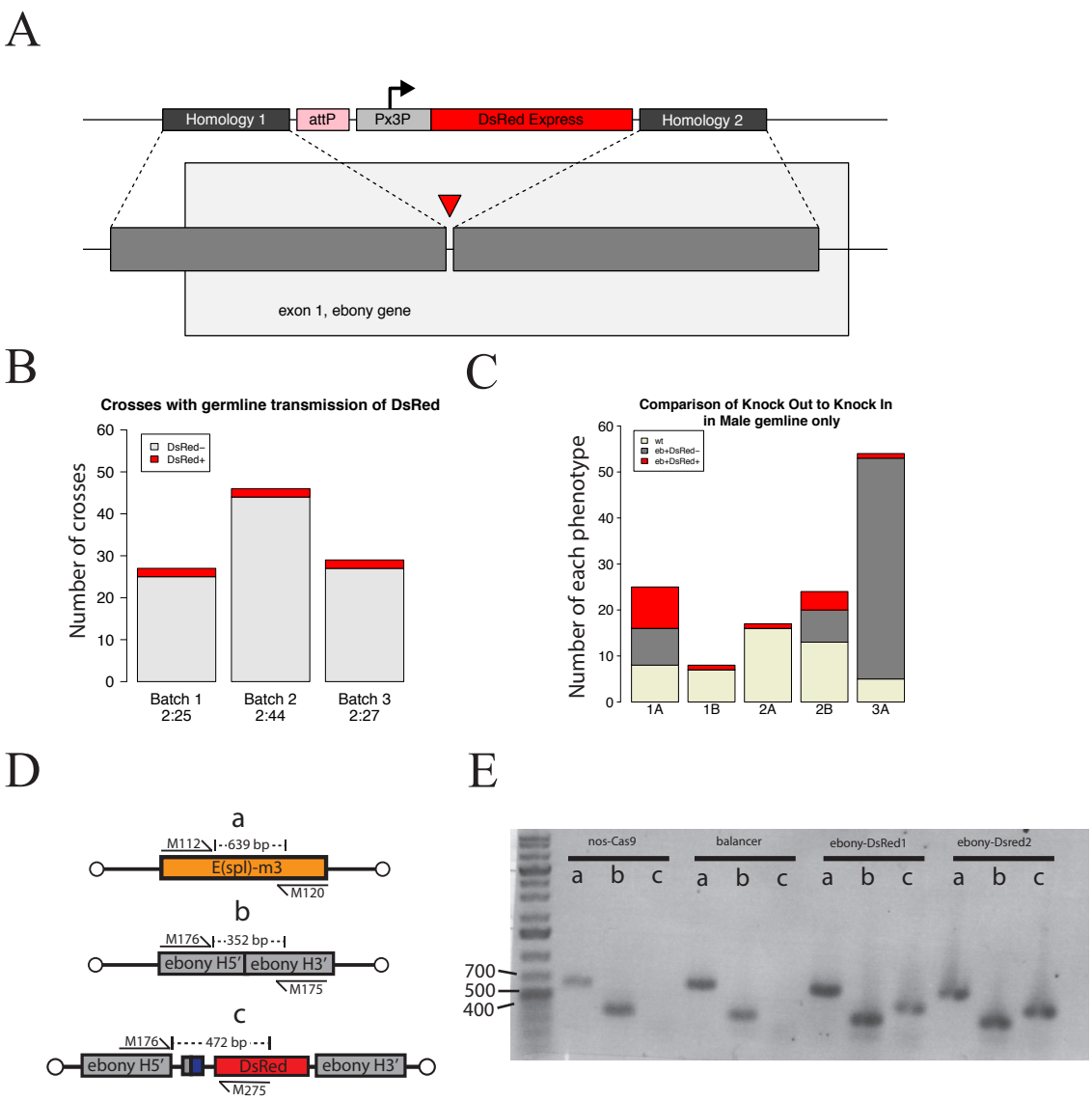


Figure 3.3: A transgene knock-in procedure using CRISPR/Cas9 at the ebony locus

Schematic of a homology directed repair construct for inducing gene knock-in at the *ebony* locus (A). Scoring of crosses with germline transmission of DsRed marker (B).Breakdown of phenotypes in 5 crosses from three separate injection batches that show germline transmission of DsRed marker (C). Schematics for PCR reactions designed to detect successful insertion of the HDR template (a and b are positive controls, presence of c indicates knock-in has occurred) (D). Genotyping gel showing amplification of positive control amplicons from parental and progeny genomic DNA and presence of ebony-DsRed fusion amplicons from progeny genomic DNA only (E).

allele and (c) an amplicon that can only be amplified from a fusion between the DsRed transgene and the sequence of the *ebony* allele 3' to gRNA site. DNA was amplified from both parental strains, pCFD2 (the Cas9 expressing stock) and w-;;TM3/TM6B (the balancer stock) to compare with the DNA from two distinct individuals both scored as carrying the DsRed marker. Genomic DNA from all strains produced amplicons from the positive control and the wild type *ebony* allele, but only those exhibiting the DsRed insertion phenotype also produced the amplicon specific to the product of the homology-directed repair (Figure 3.3E). This is the expected result given the correctly edited individuals would be heterozygous at the *ebony* locus carrying a wild type *ebony* allele on one chromosome and the DsRed insertion within the first exon of the *ebony* allele on the other chromosome.

3.2.3 Generating a platform for Int insertions at *E(spl)* via homology-directed repair.

A two-step procedure was adopted to insert the Int sequences at the locus of choice, as it was unclear whether the ParB-INT would function reliably *in vivo.* The first step was to insert attP recombination sites at the loci of interest to enable subsequent insertion of different sequences with the ΦC31 integration system (Bischof et al., 2007). In this way there would be the option to generate subsequent strains with Int1, Int2 and if these proved unsuccessful with the LacO repeats. There were also plasmids already available with the requisite sequences and a well-established pipeline for ΦC31 transgenesis in the lab.

In order to target the *E(spl)* Notch responsive locus, gRNAs were designed and constructed targeting an intergenic region at either end of this 60kb gene complex (Figure 3.4A and 3.4B). Firstly a site between *E(spl)-mδ* and *E(spl)-mgamma* was targeted and a repair template was designed to the same specification as before (Figure 3.4C). Subsequently a gRNA and repair template pair were designed to target the intergenic region at the other end of the *E(spl)* complex between *E(spl)-m7* and *E(spl)-mδ* (final construct in Figure 3.5A).

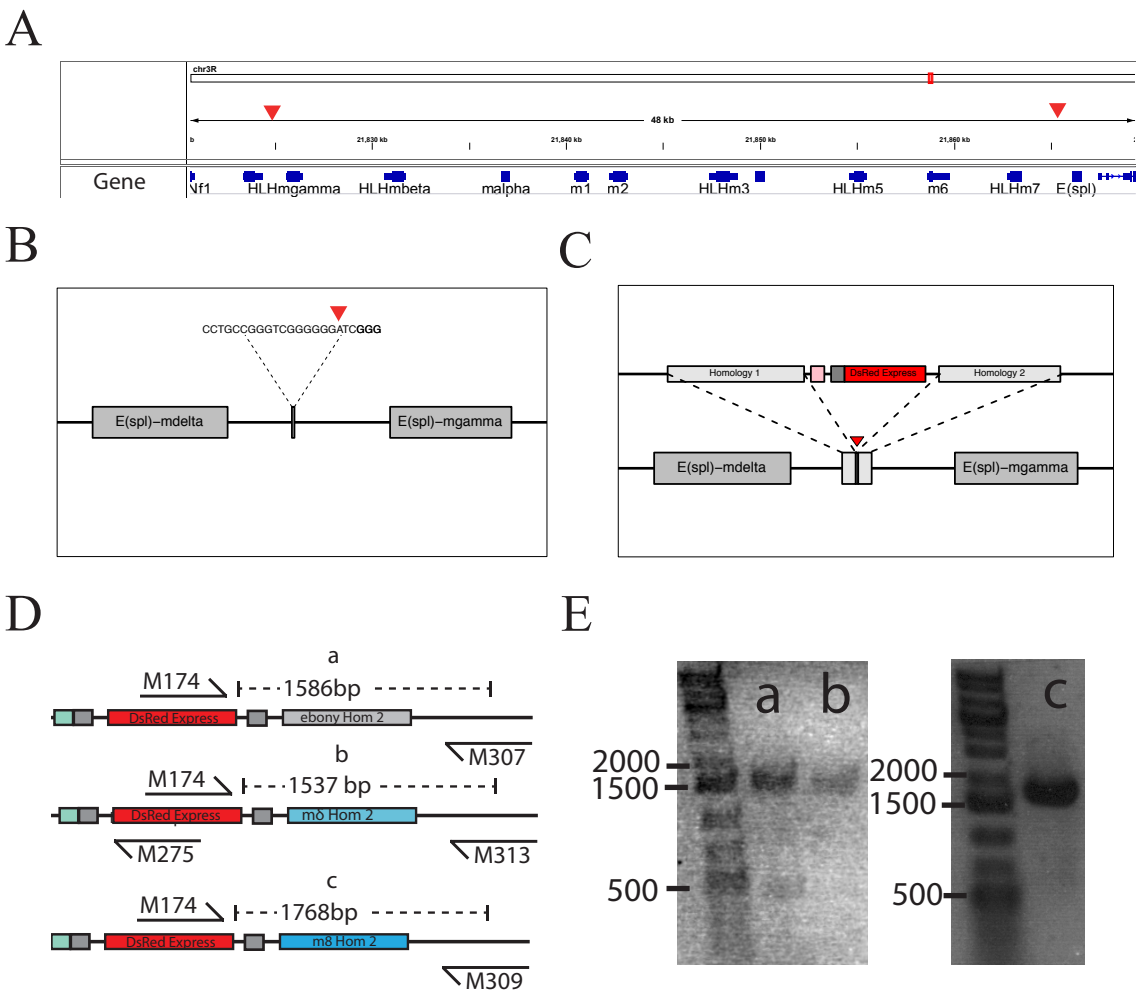


Figure 3.4: Identification of genome-edited *E(spl)* loci via PCR.

Genome browser view showing the locations of the gRNA target sites (red triangles) where the loxP-attP-DsRed-loxP constructs were inserted (A). gRNA target sequence mapping to the intergenic region between *E(spl)-mδ* and *E(spl)-mγ* , PAM sequence in bold, red triangle indicates putative cleavage site of double stranded break (DSB) 6 bases 5' of the PAM as indicated by indel location in (Cong et al., 2013) (B). Schematic showing the repair template design and mapping of the homology arms to the target locus (C). Schematics of the amplicons used to confirm presence of the DsRed marker at three HDR targeted loci (from top to bottom ebony, *E(spl)-mδ/E(spl)-mγ*, *E(spl)-m7/E(spl)-m8*) (D). Agarose gels showing positive band representing amplicons unique to successful insertions of the repair template (E).

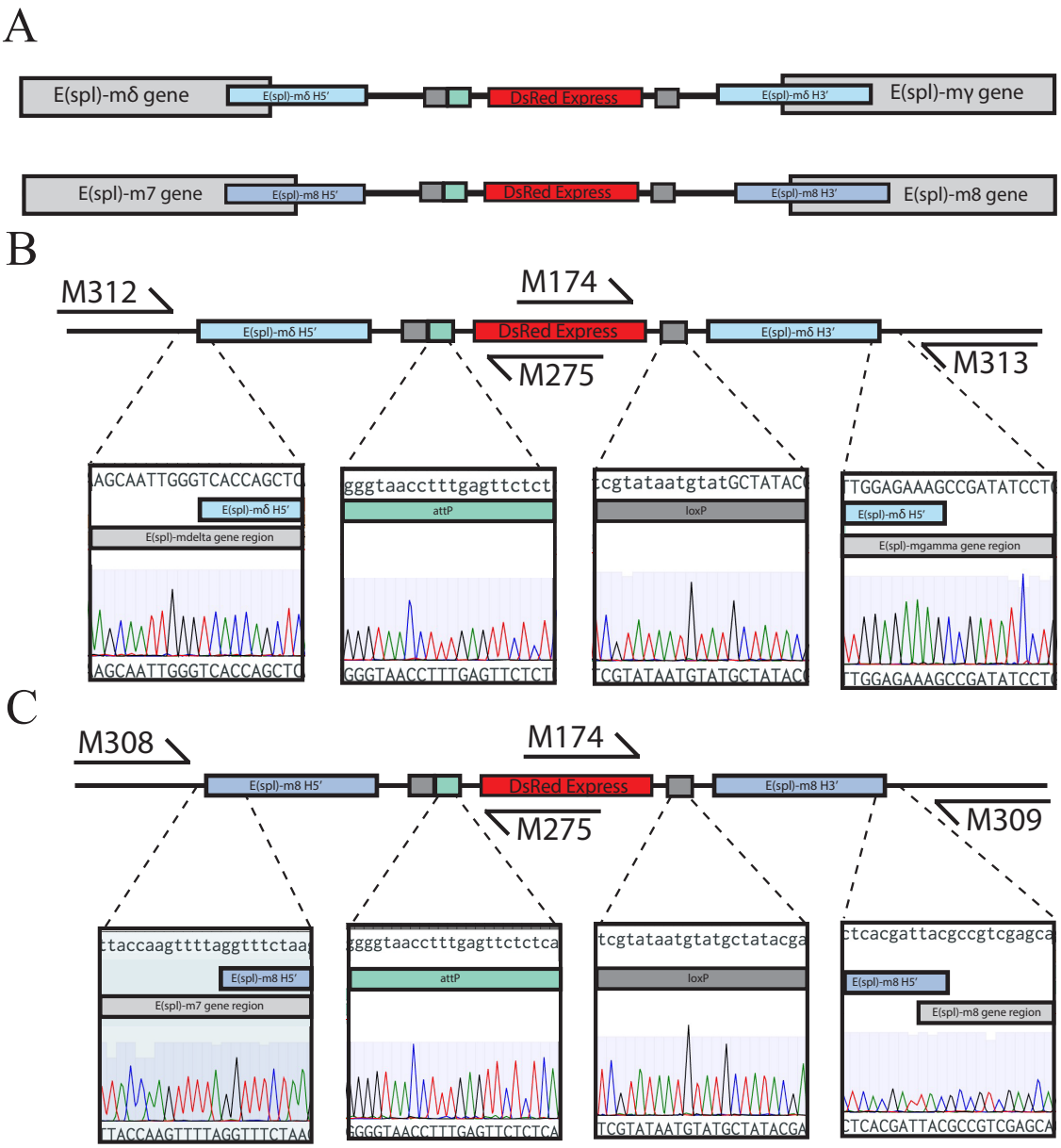


Figure 3.5: Confirmation of attP site insertion at *E(spl)* loci via sequencing.

Schematics of the constructs inserted between the genes *E(spl)-mδ/E(spl)-mgamma* and *E(spl)-m7/E(spl)-m8* (A). Excerpts from sequence traces showing key regions of the designed sequence intact. From left to right, the boundary between the *E(spl)-mδ* gene and the sequence contained within the 5' homology arm, the attP site, the 3' loxP site, the boundary between the 3' homology arm and genomic sequence not contained within the repair template plasmid (B). Sequencing trace excerpts for the insertion between *E(spl)-m7* and *E(spl)-m8* shown as for B (C). Numbered arrows indicate primer IDs and annealing region in the edited genome sequence.

Successful insertions were scored on the basis of the DsRed insertion in the male F1 progeny. The specific insertions at the *E(spl)* locus were then confirmed via PCR of amplicons specific to the desired genome edits. A similar PCR validation strategy was employed as before, using template from the first successful insertion into the *ebony* locus (a) as a positive control. DNA was successfully amplified from regions flanking the insertions as shown in the schematics in Figure 3.4D and agarose gel in Figure 3.4E. Sanger sequencing was used to confirm that successful insertions had occurred in the correct locations and orientations. Critical regions such as the junctions between the genomic sequence and homology arm sequence, and the attP sites and loxP sites were confirmed to be preserved (Figures 3.5B and 3.5C).

3.2.4 Integration of Int motifs into attP sites via ΦC31 integrase transgenesis

The second step in the locus-tagging was to insert Int1 and Int2 into the attP sites within the *E(spl)* locus. Stocks were established with the ΦC31 integrase and the relevant Int1.attB and Int2attB injected. These plasmids contain a mini-white marker allowing retrieval of transformants based on the red-eye phenotype. Once stocks had been established genomic DNA was extracted and tested for the presence of specific amplicons indicating successful integration of the plasmid at the novel *E(spl)* attP sites. As shown in the schematics (Figure 3.6A), two amplicons were chosen that extended from the 5' homology arms of the *E(spl)-mδ* and *E(spl)-m8* homology templates into sequence of the pInt1.attB and pInt2.attB plasmids. Amplicons of the predicted size were obtained in both cases (Figures 3.6B and 3.6C) indicating successful ΦC31 mediated integration of the Int-containing attB plasmids. Together with the visible markers other PCR and sequencing evidence described above, this confirmed that the strategy outlined in Figure 3.1 successfully delivered the Int1 and Int2 motifs to two bespoke insertion sites between *E(spl)-mδ/E(spl)-mgamma* and *E(spl)-m7/E(spl)-m8*.

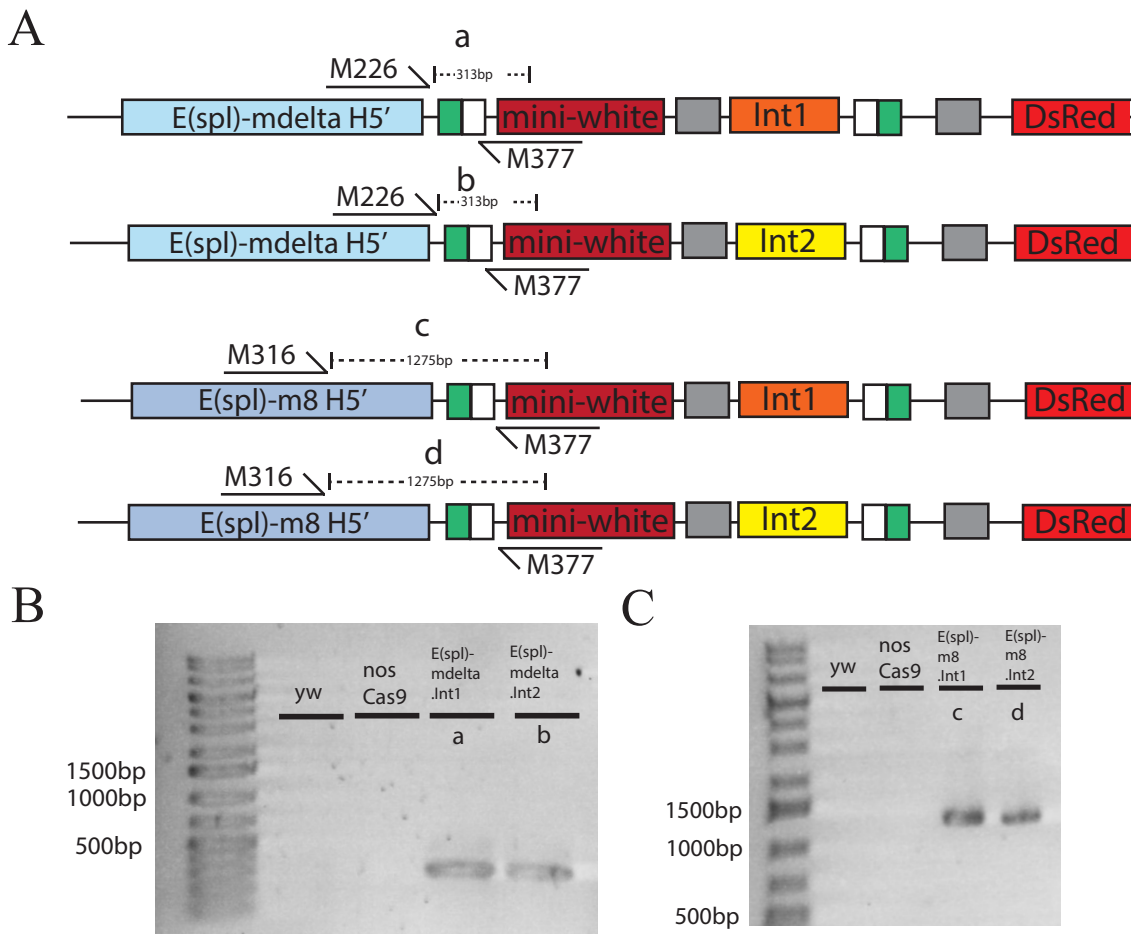


Figure 3.6: Confirmation of Int plasmid integration into *E(spl)* locus attP sites.

Schematics of the final sequences following stage two of the locus tagging transgenesis procedure outlined in Figure 1 (A). LoxP sites are shown as grey boxes, green and white boxes represent recombined attB/attP sites (attL and attR sites). Agarose DNA gel confirming that Int1 and Int2 are present in genomic DNA extracted from the *E(spl)-mδ*.Int1 stock (B). PCR amplicons specific to the junction between *E(spl)-m8*-5' homology arm and integrated pInt2.attB plasmid (C).

Comparing the results from the different loci targeted, knock in constructs for the *ebony* locus produced more fertile crosses but those targeting *E(spl)* had a higher rate of germline transmission (Figures 3.7B and 3.7C). During the establishment of the knock-in procedure the stocks were scored for viability, fertility and transmission of the desired insertion. It was noted that less embryos survived to larval stages following gRNA injections than attB plasmid injections (Kat Millen, personal communication) suggesting that genome-editing is potentially more toxic than attP/attB integration. This is possibly due to double stranded breaks induced by the Cas9 enzyme, off-target effects or sub-optimal injection concentrations. Survival was higher for the *ebony* locus, than when targeting *E(spl)* (Figure 3.7A), possibly because it is a non-essential gene, whereas targeting *E(spl)* likely results in a variety of damage to important developmental transcription factors residing in the *E(spl)* locus. In spite of the poorer survival germline transmission of DsRed was higher for the insertions into the *E(spl)* locus whether compared to the number of fertile crosses or the total number of embryos injected (Figure 3.7B and 3.7C). Finally, the proportion of successful attB plasmid integrations is recorded as a ratio (from 300 embryos injected) as this may be useful if these attP sites are used again in future to deliver plasmid cargos to the *E(spl)* locus.

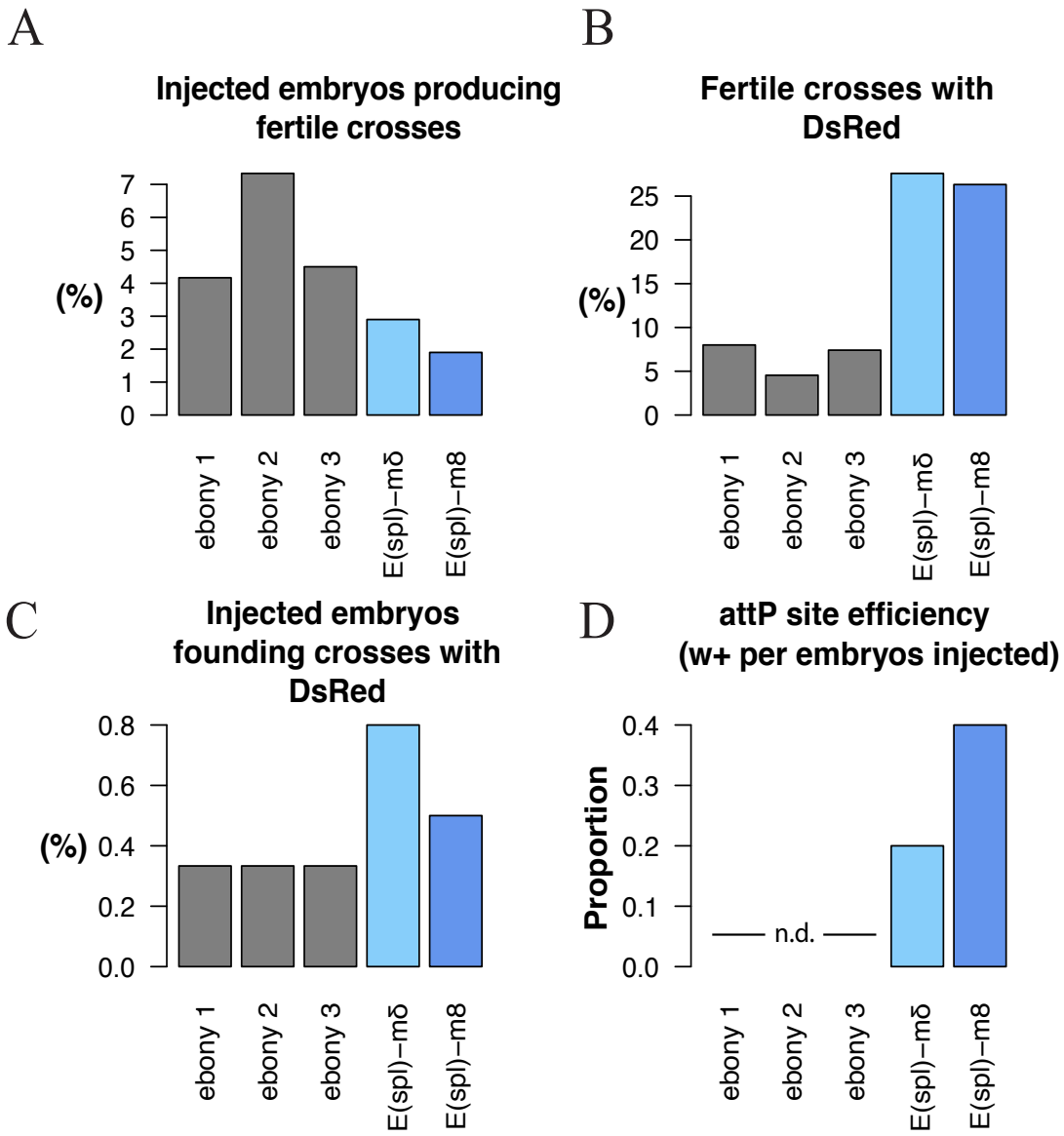


Figure 3.7: Transgenesis efficiency at different stages of locus tag generation.

Proportion of fertile crosses from founder males as a percentage of batch size (A). The number of crosses with DsRed+ progeny as a proportion of fertile crosses (B). The number of crosses with DsRed+ progeny as a proportion of embryos injected (C). Percentage of injected embryos founding crosses with white+ progeny (attP site transformation rate), not determined for ebony.attP flies (D).

3.3 Discussion

This chapter has described the establishment of a CRISPR/Cas9 genome-editing procedure that enables almost any locus of the *Drosophila* genome to be labelled with the ParB-Int locus tag system. As will be described in later chapters, these tools provide a robust and flexible method to reproducibly image the same genetic locus in combination with a wide range of imaging approaches.

The initial optimisation of CRISPR/Cas9 genome-editing based on a published gRNA sequence allowed a reliable test of the tools with an easily scorable readout of the genome-editing activity of the Cas9 enzyme expressed in the germline of the pCFD2 fly line. This strategy enabled the modification and improvement of the DNA injection concentration whilst maintaining the visualisation of the efficacy. As inserting a large construct to the ebony locus is predicted to cause the same visible body pigment as creating much smaller indels into the first exon of the *ebony* locus this starting point was an excellent stepping stone into testing homology-directed repair using constructs several kilobases long. The addition of a second visible marker at this stage again led to straight forward detection and quantification of the successful repair rate at the desired locus. In summary, the initial optimisation phase of this chapter clearly demonstrated reproducible, locus specific targeting of recombinant DNA insertions into the *Drosophila* genome. In addition to the visible scoring recorded, a PCR strategy was devised that confirmed that the repair template was inserted in the desired locus and had not been randomly integrated into the genome or remained in circular plasmid form in the progeny of the injected individuals.

3.3.1 Novel locus tags for visualising Notch responsive loci with the ParB-Int system

The groundwork establishing and optimising our knock-in procedure was then adapted to target Notch responsive loci of interest, namely two intergenic regions between *E(spl)-mδ/E(spl)-mgamma* and *E(spl)-m7/E(spl)-m8* genes in one of the best characterised Notch responsive loci. Multiple regions from the transgene insertion were

amplified by PCR and sequenced to ensure not only that the transgene was indeed located in the desired place within *E(spl)* locus, but also that all important sequence motifs such as the attP site and the loxP sites flanking the DsRed marker (which can enable subsequent removal of the marker) had been delivered to the insertion site intact. After successful tagging of the *E(spl)* loci, essentially the same procedure was followed to tag two more loci: *Hey* with Int1 and *dpn* with Int2. These loci are in the following chapters (genotyping details for *Hey.IntA* and *dpn*.Int2 in Appendix 2).

The tools generated here are likely to be widely reusable for a variety of different experiments. Gene specific locus tags compatible with immunostaining and live imaging will hopefully mean future researchers will be able to study the Notch responsive loci of *Drosophila* with single gene resolution in a way that enables a comparison between a single locus and the rest of the nucleus or genome. As suggested earlier, these insertions can be paired to visualise multiple loci simultaneously, and thus reveal how genetic loci interact in 3D space. The novel attP sites generated in E(spl) could also potentially enable further modification of the E(spl) locus exploiting the backwards compatibility with existing attB vectors. Perhaps most exciting, is the possibility for combining this chromosome labelling system with *in vivo* imaging to study the dynamics of genetic loci and the proteins that interact with them.

3.3.2 Off-target effects of CRISPR/Cas9 genome-editing

Although CRISPR/Cas9 genome-editing is an incredibly powerful tool there are number of important caveats that can undermine studies which use this approach. Due to the fact that this genome-editing approach relies on the homology between a gRNA and the genome, the accuracy of the Cas9 nuclease activity is only as good as the specificity of the gRNA. Off-target binding and cleavage have been documented in a number of studies using CRISPR/Cas9 and the genome-editing field is already researching ways to control and minimise them (Ren et al., 2014). In order to minimise the chance of off-target effects we used a published computational resource with the maximum stringency settings to filter out gRNA sequences with likely off-target effects and also only selected gRNA sequences with the on-target efficiency of 6 or higher as recommended (see methods)(Ren et al., 2013). In the future, if the cost of whole genome sequencing continues to drop it may become standard practice to screen transformants for

problematic off-target mutations, however in the absence whole genome sequencing, off target effects cannot be ruled out.

3.3.3 Potential improvements to the genome-editing procedure

Since the genome-editing experiments in this study were carried out evidence has arisen that linear templates for homology-directed repair may be more efficient than circular plasmid templates, so future HDR experiments may be improved by making this small alteration to the protocol either by digesting the plasmid at a unique site in the backbone or injecting linear PCR amplified homology templates (Song and Stieger, 2017). The two step approach for locus tag insertion required long-term transgenesis and had low efficiency at both stages. However, it did allow the flexibility to insert LacO repeat arrays into the same locus as an alternative should the ParB-Int system have failed in any way. The novel attP sites generated in this study could in fact be labelled with LacO repeats and used in conjunction with the ParB-Int system to label three loci simultaneously, though ideally 3 different fluorescent proteins with non-overlapping spectra would be required for easiest detection. Future studies could produce locus tags at additional loci in one step using the ParB-Int system, this would be especially straight forward if the Int motifs are combined with markers such as the DsRed marker used in this study or similar. Recent advances in genome-editing in the majority of popular model organisms will undoubtedly liberate locus tagging tools to be more widely used in the future.

3.3.4 Conclusion

The advent of rapid and efficient genome-editing in *Drosophila* has already begun to provide many exciting opportunities for interrogating the genome of this already well characterised model organism. By combining this powerful tool with one of the most recently developed chromosomal labelling strategies it is hoped that the location and dynamics of Notch responsive loci can be revealed. As will be described later in this

thesis, these tools have already begun to reveal the subnuclear localisation of Notch responsive genes.

4 Quantification of subnuclear gene position in salivary gland nuclei

4.1 Introduction

The salivary gland of the *Drosophila* third instar larva is an excellent model system in which to study the nucleus and the organisation of the genome due to the polytene chromosomes and relatively large size of the nuclei. Moreover, the mature, post-mitotic secretory endothelial cells that make up the majority of the gland's volume have no detectable level of active Notch signalling. Together, these properties of the salivary gland nuclei present the opportunity to image nuclear organisation in great detail using the tools described in the previous chapter, and to introduce an ectopic Notch signal in order to determine any effects driven by the addition of Notch to cells that would otherwise not receive any Notch signal at the stage of development in question.

Polytene chromosomes have long been a popular system in which to visualise chromosomal loci in part due to the striking banding pattern which has been used to generate detailed cytogenetic maps of the *Drosophila* genome. Mapping of polytene

chromosomes has primarily relied upon the so-called 'chromosome squash' method which is still commonly used to interrogate the chromosomes with immunolabelling or in situ hybridisation approaches (Zielke et al., 2016). This method involves flattening the nucleus and spreading the DNA, usually between two glass slides, providing a two-dimensional chromosome map but omitting the three-dimensional positional information regarding organisation of chromosomes in the nucleus. Thus the subnuclear localisation of individual chromosomal loci in *Drosophila* salivary gland nuclei has not been well studied with a nucleus-by-nucleus, imaging based approach. A detailed physical mapping study of the arrangement of polytene chromsomes has previously concluded that the intranuclear position of genes in the salivary gland nucleus is not important for their regulation (Mathog and Sedat, 1989). However, this study did not involve labelling individual genes or placing this information in the context of gene expression analysis.

More recently the application of high-throughput chromatin-conformation-capture techniques (Hi-C) to the salivary gland chromosomes has shown that TADs revealed in resulting contact maps correspond to the banding patterns on polytene chromosomes. In this study 95% of polytene bands corresponded to TADs in the polytene chromosome Hi-C maps. Moreover, TADs in salivary gland nuclei correspond to TADs detected by Hi-C in diploid cells more often that would be expected due to chance (Eagen et al., 2015). Over 50% of polytene TADs are within 40kb of a TAD in diploid cells implying that the gross organisation of the genome in 3D space is remarkably similar between the cell types compared, possibly indicating that a substantial portion of genome organisation is conserved between highly differentiated cell types. Whilst Hi-C and related approaches provide a wealth of information about population level contact frequencies between different genome regions they give little insight into the localisation of individual genes within a single nucleus. It remains unclear how gene position varies from cell to cell, in a given nucleus over time or if it varies before and after a signalling event.

In order to study the subnuclear localisation of Notch responsive genes, four novel locus tag insertions were generated as described in the previous chapter (the genomic locations of these locus tags and controls shown in Figure 4.1). In addition to the

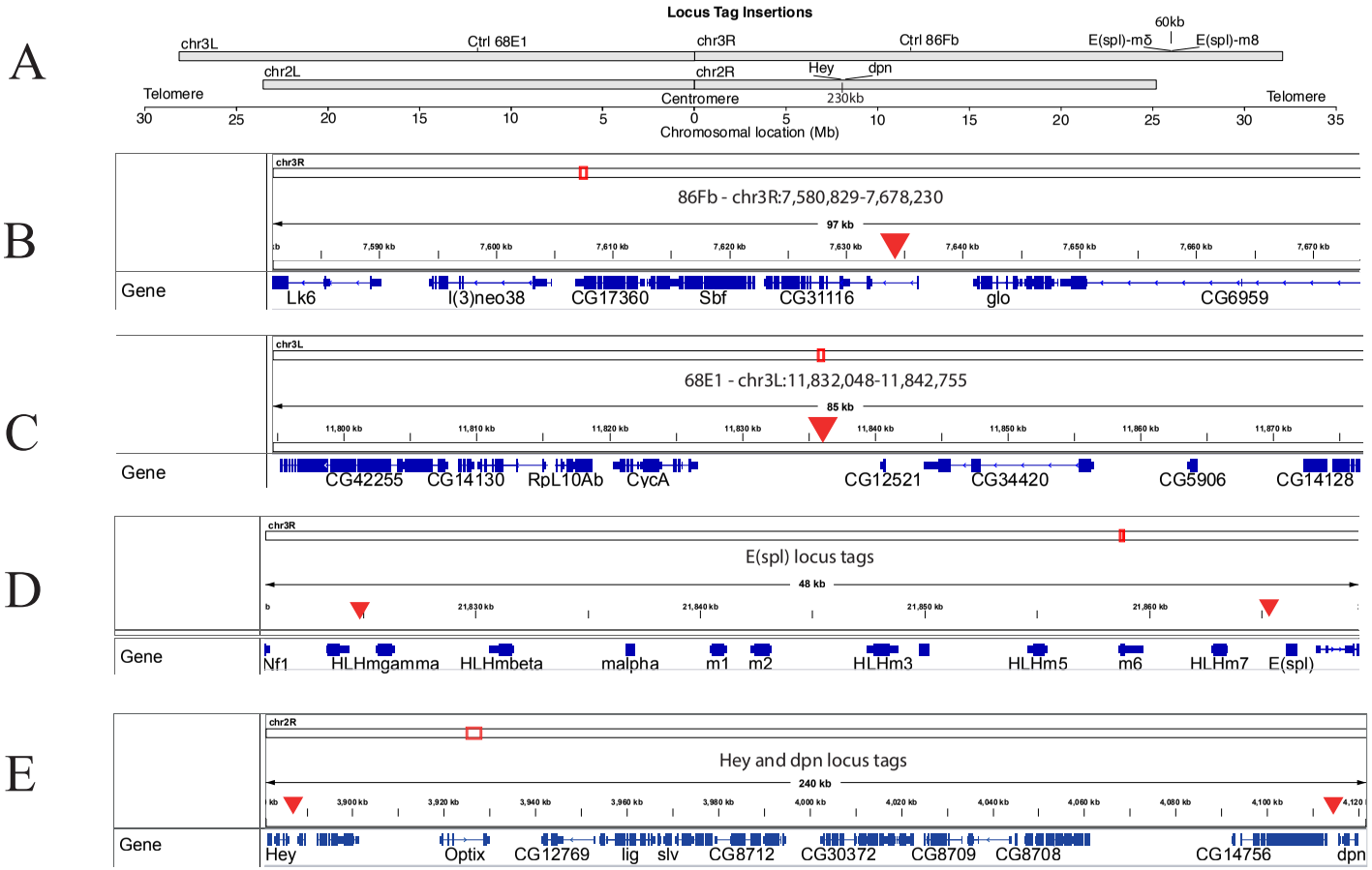


Figure 4.1: Overview of the locus tag positions used in this study.

Schematic of 6 locus tag insertion points on *Drosophila* autosomes 2 and 3. Ctrl-86Fb (B) and Ctrl-68E1 (C) were generated by collaborators using standard ΦC31 integrase transgenesis at existing attP sites. The remaining six locus tags are *E(spl)-mδ*.Int1, *E(spl)-mδ*.Int2, *E(spl)-m8*.Int1, *E(spl) m8*.Int2 (D) and *Hey*.Int1, *dpn*.Int2 (E) were tagged via the two step process described in chapter 3, using attP sites generated in this study with intergenic regions in *E(spl)-C*

insertions at *E(spl)-mδ*/*E(spl)-mγ* and *E(spl)-m7/E(spl)-m8* intergenic locations, described in chapter 3, the locus tagging procedure was repeated for intergenic regions close to *Hey* and *dpn* (see Appendix 2 for genotyping data for *Hey*.Int1 and *dpn*.Int2). In all cases the new insertions were made into intergenic regions within 3kb of the chosen Notch target. The *Enhancer of split* (*E(spl)*) locus is a very well characterised Notch responsive locus approximately 50kb long with a high density of Notch responding genes. In cultured Kc167 cells the locus responds to an artificial stimulation of the Notch pathway within 10 minutes as detected by qPCR or microarray studies. Gene expression can change as much as 50 fold or more in just 30 minutes in the case of

E(spl)-m3 (Krejci et al., 2009). The chromatin state of these genome regions is known from DamID studies and both *E(spl)* and *dpn* are found within regions of 'blue' chromatin characterized by widespread binding of polycomb proteins and H3K27 trimethylation. In contrast *Hey* is found in 'black' chromatin in which large amounts of histone H1, d1 and lamin are detected. Black chromatin is the most prevalent chromatin state and houses the majority of silent genes and many with very low expression levels (Filion et al., 2010). Although it is unknown whether the tagged loci are able to respond to Notch in the salivary gland, *E(spl)* was likely to respond based on its expression in other tissues. The majority of the genes neighbouring the locus tags have very low or undetectable expression in the salivary gland based on the modEncode anatomy RNA-seq dataset (Figure 4.2).

Whilst Notch signalling is known to be involved in maintaining the stem cells in many tissue types and is required in *Drosophila* for normal salivary gland development (Dang et al., 2009; Hartenstein et al., 1992), there is no evidence that Notch signalling is active in the mature cells of the salivary gland. In agreement the expression levels of the Notch receptor and ligands Delta and Serrate are recorded as 0 in the modEncode Anatomy RNA-seq data set. Additionally fluorescence from the synthetic Notch activity reporter 'NRE-GFP' is observed only in ring cells around the salivary gland duct (Gomez-Lamarca et al. manuscript under review). Therefore, the gene positions observed in the nuclei of mature glands will correspond to the Notch OFF state. Furthermore, this state can be altered experimentally by supplying a constitutively active truncated Notch receptor under control of the Gal4/UAS system.

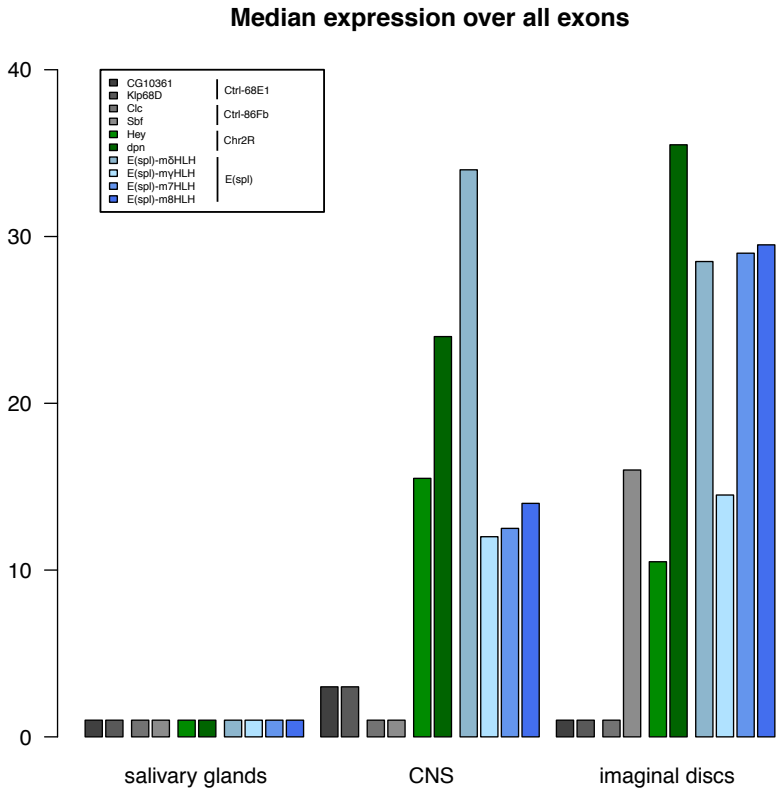


Figure 4.2: Expression of genes neighbouring the two control and four Notch-responsive loci. Data from flybase, modEncode Anatomy expression dataset.

The median expression across all introns is plotted for the two tissues addressed in this study, salivary gland and CNS, and imaginal discs, which also have some areas of high Notch signalling, as a comparison.

The aims of the experiments in this chapter were to confirm the specificity of the ParB-INT locus tag system in labelling the tagged regions associated with the genes of interest, to determine the distribution of gene locations in the nucleus so that any stereotyped positions would be revealed and to investigate whether the localisation of a Notch-responsive gene changes under ectopic Notch activation. Two quantification methods are discussed, one using absolute distances within 3D confocal stacks, the latter using relative distance maps based on an eroded volume fraction (EVF) approach. The latter is then used to determine the relative radial position of each of the tagged genes and the change in position of the *E(spl)-m8* locus tag following a strong, ectopic Notch activation.

4.2 Results

4.2.1 Specificity of the locus tags

The ParB-INT locus tagging system relies on the sequence-specific binding of the ParB1 and ParB2 proteins to their respective motifs, Int1 and Int2. The specific localisation of the ParB proteins was initially tested using Int motifs inserted into control loci. The schematic in Figure 4.3A indicates the basic principles of the ParB-Int system and Figure 4.3B shows examples of salivary gland nuclei. Single bright fluorescence foci were detected on polytene chromosomes when the Int locus was paired with its cognate ParB protein. Such foci were absent from nuclei in individuals bearing the non-cognate pairs (e.g. ParB1-mCherry / Ctrl.Int2).

Similar controls were performed for the genome-edited *E(spl)-mδ.Int1* and *E(spl)-m8.Int1* lines to confirm that the ParB1-mCherry signal was specifically accumulating at the tagged locus. Firstly, salivary gland nuclei with locus tag cognate pairs were imaged alongside controls bearing a wild type chromosome from crossing ParB1-mCherry flies to *yw* males. This confirmed there was no significant non-specific ParB1-mCherry accumulation in the absence of an INT sequence insertion (Figure 4.4A, top). The presence of bright fluorescent foci in the crosses using the novel transgenics bearing insertions to the *E(spl)* loci (Figure 4.4A middle and bottom) confirmed that a signal was observed in the same manner to that seen using the control locus tags (Figure 4.3B).

To further investigate the specificity of the signal to the E(spl) locus, recombinants were generated containing a locus tag and a cognate Par on the same chromosome. Crossing flies bearing ParB1-mCherry ~ *E(spl)-m8.Int1* with ParB2-GFP ~ *E(spl)-mδ.Int2* produced progeny with two recombinant third chromosomes allowing the visualisation of the two *E(spl)* locus tags simultaneously (genetics and imaging for Figure 4.4B by Zoe Pillidge). As predicted for tags located within 50kb of one another the locus tag signal in the red and green channels was juxtaposed. This confirms that two locus tags label

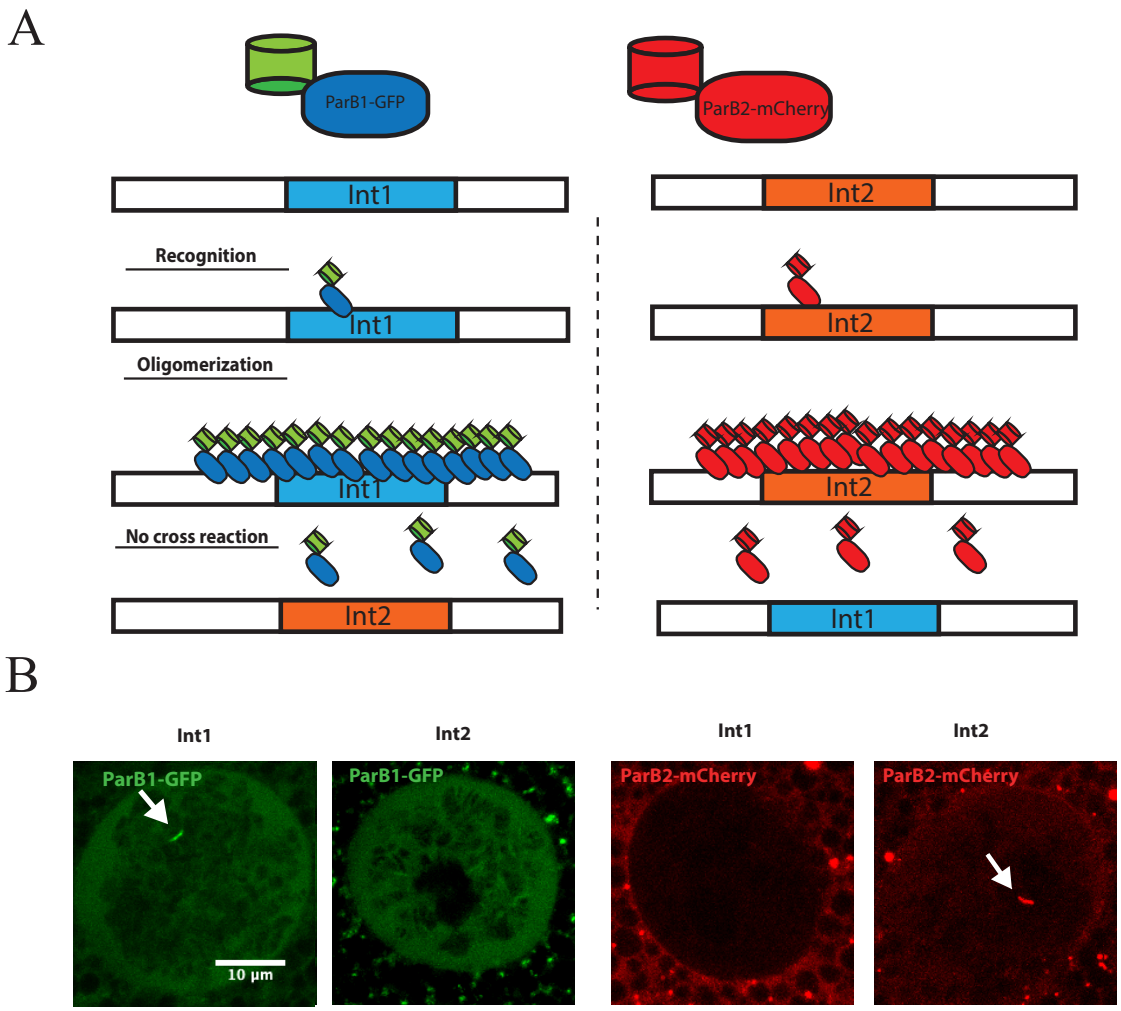


Figure 4.3: The ParB-Int system labels loci by oligomerisation and comprises two non-cross reacting pairs of DNA-binding proteins and corresponding binding motifs.

Schematic showing the recognition and auto-recruitment as described in (Saad et al., 2014) (A). Crossing fly strains expressing ParB1 to those bearing Int1 produces visible foci, as suggested in the original publication whereas crossing in the Int2 motif instead does not. The reciprocal crosses test the specificity of ParB2 proteins for Int2 sequence and lack of reactivity with Int1 sequence. Examples of nuclei from ParB1-GFP (left) and ParB2-mCherry (right) salivary gland nuclei (B). Scale bar 10 microns, all samples were imaged under identical magnification.

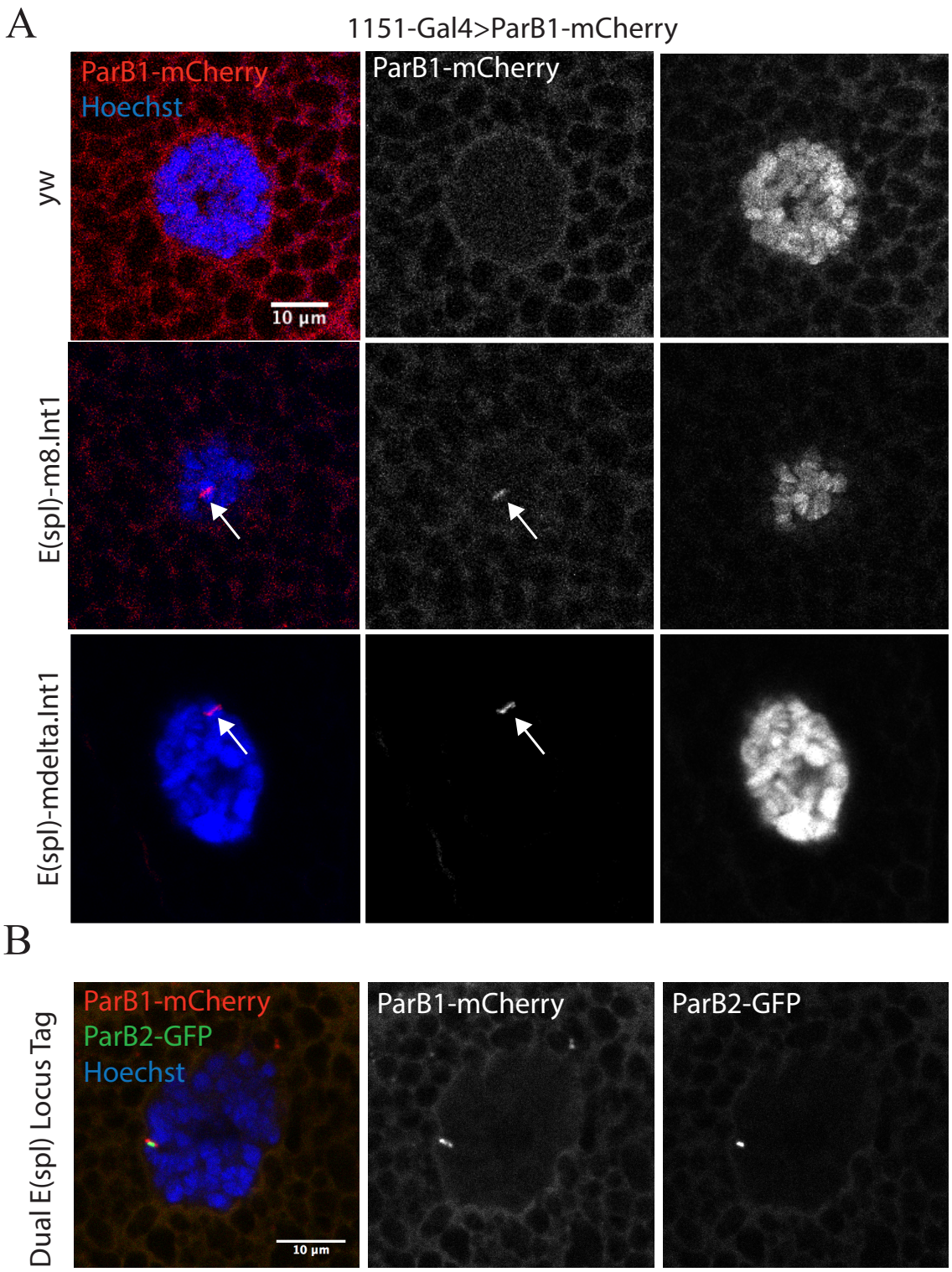


Figure 4.4: Examples of specificity controls for the locus tags at *E(spl)*.

Salivary gland nuclei in ParB1-mCherry expressing cells with wild type third chromosomes (negative control, top), *E(spl)-mδ*.Int1 (middle) and *E(spl)-m8* (bottom) (A). Colocalisation of *E(spl)-m8* and *E(spl)-mδ* to the same region of the polytene chromosome confirms accuracy of the genome editing described in the previous chapter and the expected functioning of the locus tag system (B). All images in Figure 4B were provided by Zoe Pillidge.

the same region of the genome (Figure 4.4B). Interestingly the two locus tags in close proximity are not precisely co-localized (no yellow signal when merged), indicating that they overlap closely in space but are bound to distinct positions. Their arrangement suggests that the paired polytene chromosomes from the male and female parent are interdigitated. Based on results from all of the preliminary experiments the locus tags are highly specific and allow detection of genomic loci *in vivo*.

4.2.2 Methods for measuring locus positions within the nucleus

To enable the position of the tagged loci to be quantified, each salivary gland nucleus was imaged in its entirety. This allowed the dimensions of each nucleus to be determined and thus each measurement to be scaled according to a radial measurement of the nucleus, which was important because a range of sizes and morphologies of nuclei were observed. As described in more detail below a raw measurement of the radial position of the locus tag was scaled by a factor related to the size and shape of the nucleus to produce a normalised position. Variations in the position were evident when the 3D images were analysed. Figure 4.5A shows two different examples of salivary gland nuclei from several different orthogonal views alongside a 3D reconstruction. This highlights the benefit of 3D quantification versus 2D quantification when considering the nuclear position of a gene and also demonstrates the irregular but roughly spherical shape of a salivary gland nucleus. 3D reconstructed ball and stick models represent the centroids from the detected nuclear boundary and the detected locus tag (Figure 4.5B, large blue sphere and small red sphere respectively). The red, blue and green lines show the distances between the points that were calculated as described below.

The first approach to quantifying the position of the locus tag in a reproducible manner was to determine the absolute distances between the nuclear centroid, the locus tag, and the nuclear periphery. Centroid coordinates for the chromosome outline and the locus tag as well as the co-ordinates of all the points that lie on the surface of the detected object were generated by the segmentation and detection pipeline (described in the methods). The distances between these points can then be calculated by applying Pythagoras' theorem, which determines the length of the longest side of a right angled

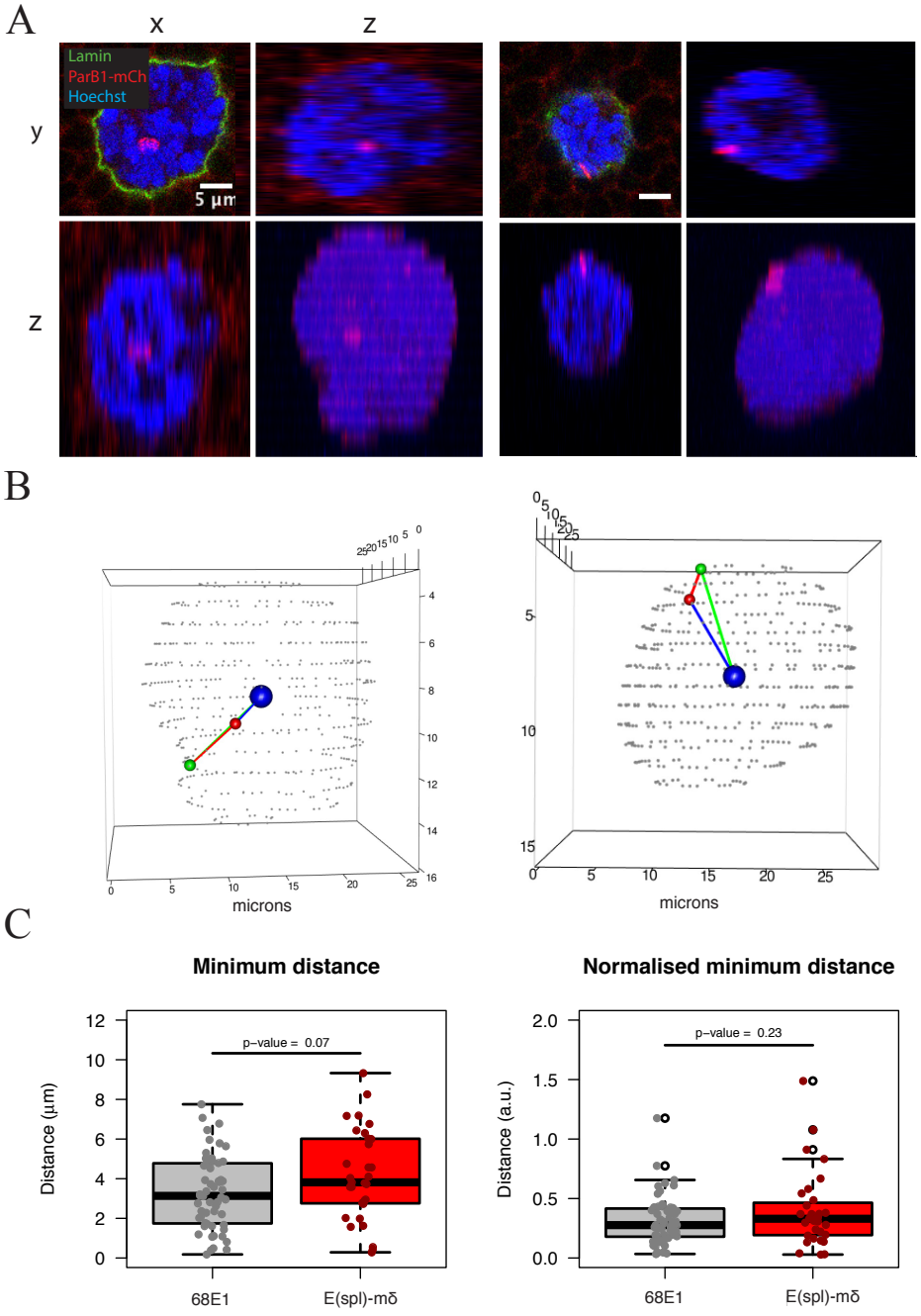


Figure 4.5: A locus tag quantification approach based on absolute measurements of nuclear dimensions.

Orthogonal views of nuclei with locus tag in the mid section (left) and periphery (right) Bottom right image is a 3D reconstruction (A) 3D skeletons of locus tag containing nuclei showing the nuclear centroid (blue), locus tag centroid (red) and locus tag nearest surface point (green) (B). Lines show nuclear centroid - locus tag centroid distance (blue), locus tag to surface distance (red) and centroid to surface distance (green). Boxplot showing raw values from minimum distance calculations and minimum distances normalised as a ratio of a nuclear radius measurement (C). White circles on boxplots indicate outliers, in this case the normalised minimum distance of locus tag positions in nuclei of low sphericity.

triangle, in three dimensions. In this case, the distance calculated is that between two points in 3D space taking into account the width, height and depth of the cubic space that separates them. Where a is the width, b is the height and c is the depth of the cubic space that separates the two points in 3D space.

Pythagoras' theorem for three-dimensional distances:

$$x = \sqrt{a^2 + b^2 + c^2}$$

This formula was applied to find the distance between the locus tag and the nuclear centroid. The same formula was applied to find the distance between the locus tag and the closest surface voxel and the distance between the nuclear centroid and the closest surface voxel to the locus tag. These distances are highlighted in Figure 4.5B as the blue line, the red line and the green line respectively. These measurements were then used to determine the locus tag to surface distance in a number of different nuclei for the locus tag positioned at 68E1.attP and *E(spl)-mδ*.attP (Figure 4.5C). Based on the raw measurements there is considerable variability in the distance between the locus tag and the periphery. The range was similar for both (between 0 and 10 microns). Given the nuclei are typically 20-30 microns in diameter this indicates that locus tags can be found at the majority of possible radial positions within the nucleus. There is a small difference in the mean value as indicated by the boxplot but this was not found to be statistically significant (t-test raw data p value = 0.07, normalised data p value = 0.22). It is possible that *E(spl)-mδ* has a more central position on average but that this experiment and analysis did not have sufficient statistical power to detect the small effect size seen in a very variable position of two chromosomal tags located at different positions on the same chromosome. However, the normalised data shows a reduction in the already small difference between the mean position of the two loci suggesting that any differences may be due to the size of the nuclei differing between the two samples. Specifically, if the nuclei imaged with the *E(spl)-mδ* locus tag were slightly larger on average, then there would be more chance of raw measurements having a longer locus tag to surface distance than the control.

Formula for normalising the locus tag to surface distance to the radial distance and the nearest surface point:

$$normalised\ distance = \frac{raw\ locus\ tag\ to\ surface\ distance}{radial\ length\ at\ locus\ tag\ nearest\ point}$$

This initial attempt at quantifying locus tag position highlighted a key feature of the data that drove further development of an improved quantification method. Firstly, the images in Figure 4.5A and 4.5B show that the salivary gland nuclei are not perfectly spherical and a simple geometric quantification method may not account for the irregularities in their 3D shape. This is indicated by a number of outlier values in the normalised data versus the raw values in Figure 4.5C. Moreover, as pointed out earlier, the differences in the raw values could have been accounted for by variation in nucleus size. Thus a method which could account for irregularly shaped nuclei, independent of size was sought.

One approach that has been used in 3D distance analysis is to divide an object into shells of equal volume and score each point based on the shell it falls into, such as 0 for the outermost shell and 1 for the inner most shell and intermediate shells at 0.25, 0.5, and 0.75 for example. By using shells of equal volume one can easily determine if the positioning of the object in question is non-random, as a randomly positioned object would appear in each shell with equal frequency if sampled enough times. An extension of this approach to a continuous scale is known as eroded volume fraction (EVF) (Ballester et al., 2008). This approach was implemented in FIJI by using a Euclidean distance transform on the binary mask of the nuclear outline. This transform converts each white pixel in the starting mask into a shade of grey representing its relative distance from the centre.

The results of the eroded volume fraction approach for two nuclei is illustrated in Figure 4.6. The corresponding detection outlines and distance maps are shown in the top two rows (red outlines show 3D segmentation detection outline, top, grayscale EVF maps, middle). Each shade of grey in the EVF maps indicates a pixel is assigned to a shell of

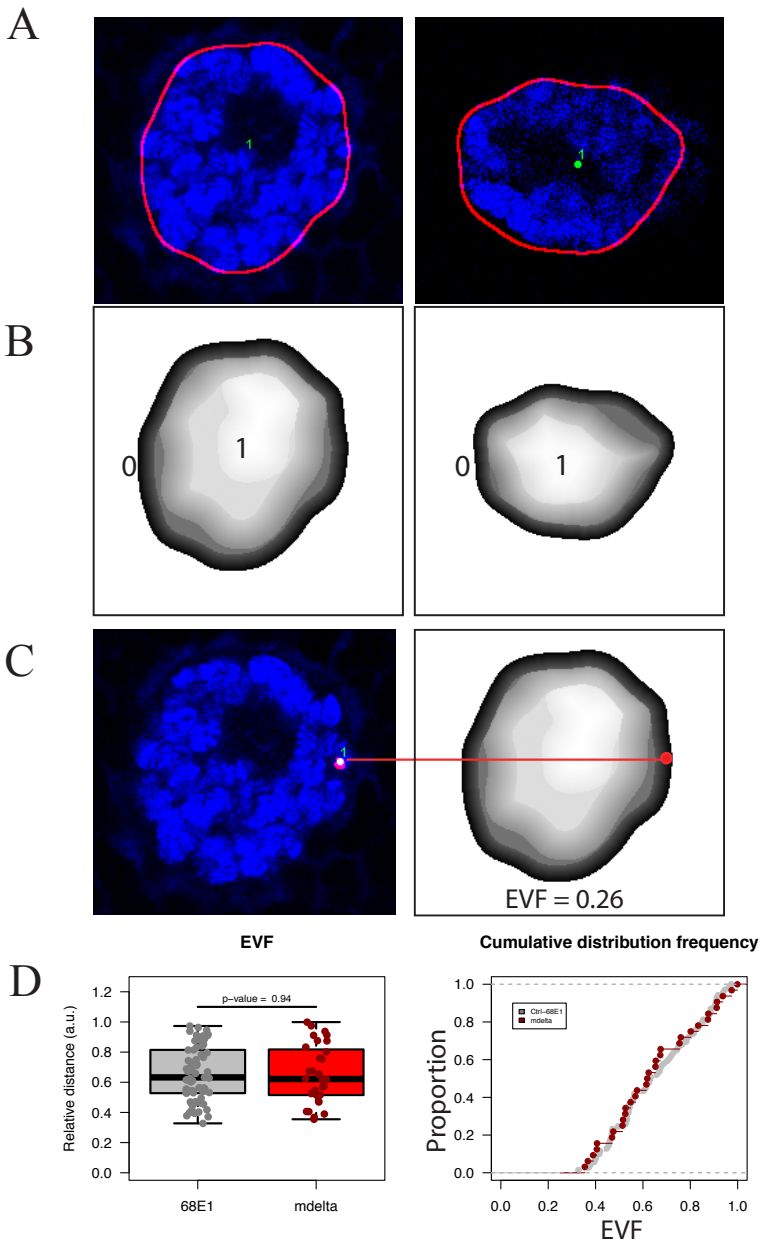


Figure 4.6: Nuclear perimeter detections converted into a volume based map of relative radial distance.

Two nuclei showing extremes of 3D shape, the nucleus on the left has high sphericity value (>0.6) and the one on the right has lower sphericity (<0.25) (A). Eroded volume fraction based distance maps shaded to show concentric shells of equal volume (B). The gradient spans from 0 (black) in the periphery to 1 (white) in the center. A locus tag detection determines 3D co-ordinates for the locus tag centroid, the EVF score is obtained by determining the pixel value in the 3D distance map (C). Boxplots show EVF values for the same nuclei and locus tags quantified in Figure 4.4 (D). On the right the empirical cumulative distribution frequency is plotted to show the distribution of the relative radial distances for each locus tag, (two-sample Kolmogorov-Smirnov test p-value = 0.94).

equal volume across a continuous scale from periphery (0, black) to center (1, white). .

In order to determine the EVF score of a given locus tag detection, the pixel value is looked up on the EVF map (Figure 4.6C). This approach gives a normalised read out of radial position (Figure 4.6D). Furthermore, when all the nuclei were analysed there were no obvious outliers, indicating that this normalisation method is robust in the face of the nuclear volume variation that caused a difference in the raw values in Figure 4.5C. The same values can be plotted as an empirical cumulative frequency distribution (Figure 6D, right) which shows that the positions from both locus tags fall in almost identical distributions. Comparison of the EVF values measured for the locus tag at 68E1.attP and *E(spl)-mδ*.attP with a Kolmogorov-Smirnov test found no significant difference (p value = 0.9018). Given that the two different quantification methods resulted in the same conclusion that there is very little difference between the localisation of the locus tag at 68E1.attP and *E(spl)-mδ*.attP it is likely that the localisation of the locus tags is determined largely at random or a number of variable factors that produces a distribution of positions that varies randomly. Except however for one key difference between the absolute distance method in Figure 4.5 and the EVF method depicted in Figure 4.6 which is that the locus tag appears to be excluded from the outer 30% of the nuclear volume, though this effect is likely to be a result of the thickness of the polytene chromosomes, because even a peripherally located locus tag (Figure 4.6C) has an EVF value of 0.46, indicating that there is a substantial volume (30%) outside the centroid of this locus tag which would correspond to a portion of the chromosomes.

4.2.3 Quantification of relative locus tag position in 3D using eroded volume fraction

The eroded volume fraction approach was then applied to each locus tag in turn, aiming for n of more than 30 nuclei in each case. A cursory analysis suggested that the *dpn*.Int2 locus tag was commonly found in the extreme periphery of the nucleus. This was

verified when the segmentation pipeline and EVF quantification method were applied to this dataset. The *dpn*.Int2 locus tag had a strikingly different distribution than the control locus at 68E1.attP (Figure 4.7A). Whilst the Ctrl.68E1 locus tag is distributed evenly between 0.3 and 1 on the EVF scale, *dpn*.Int2 is primarly positioned below 0.3 and between 0.5 and 0.7. A two-sample Kolmogorov-Smirnov that the localisation of the locus tags at 68E1.attP and *dpn*.attP are distinct in a manner that is highly unlikely to be due to random chance (p-value <0.001). Furthermore, the distinct positioning of the *dpn* locus tag at either the periphery or the nuclear centre may indicate a strong association between *dpn* and zones of heterochromatin and nuclear landmarks such as the nuclear envelope and encircling the nucleolus.

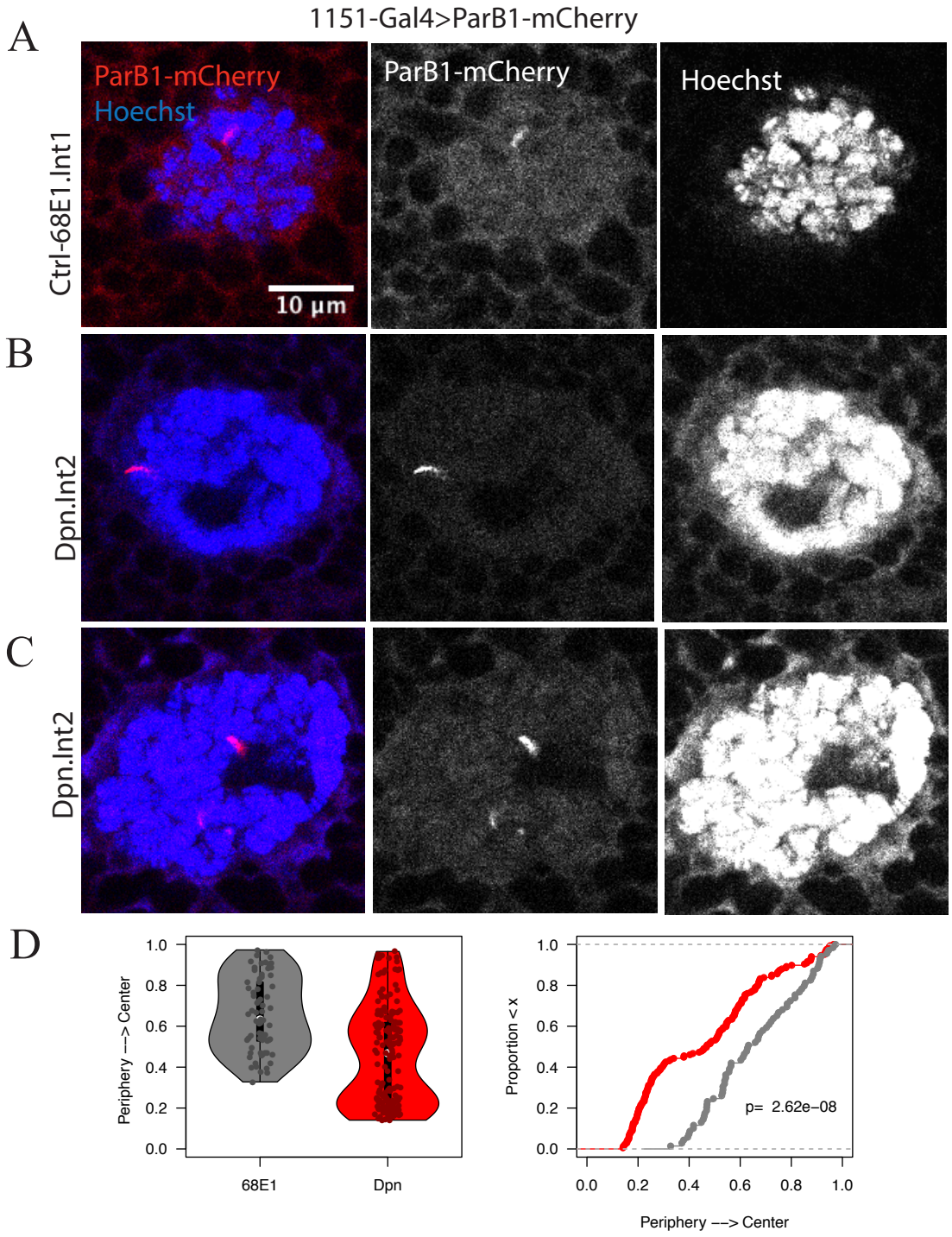


Figure 4.7: *dpn* is observed at the nuclear periphery and the nucleolus.

An example of the 68E1.Int1 locus tag visualised in ParB1-mCherry-expressing salivary gland cell (A). This is the most peripheral example of this locus tag detected. Example of *dpn*.Int2 locus tag at the nulear periphery (B). Example of the *dpn.*Int2 at the nucleolus (C). Violin and scatter plot showing distribution of Ctrl-68E1.Int1 and *dpn.*Int2 locus tags across the EVF scale (left). Cumulative distribution plot of the same distributions, two-sample Kolmogorov-Smirnov test, p-value = 2.62×10^{-8} (D).

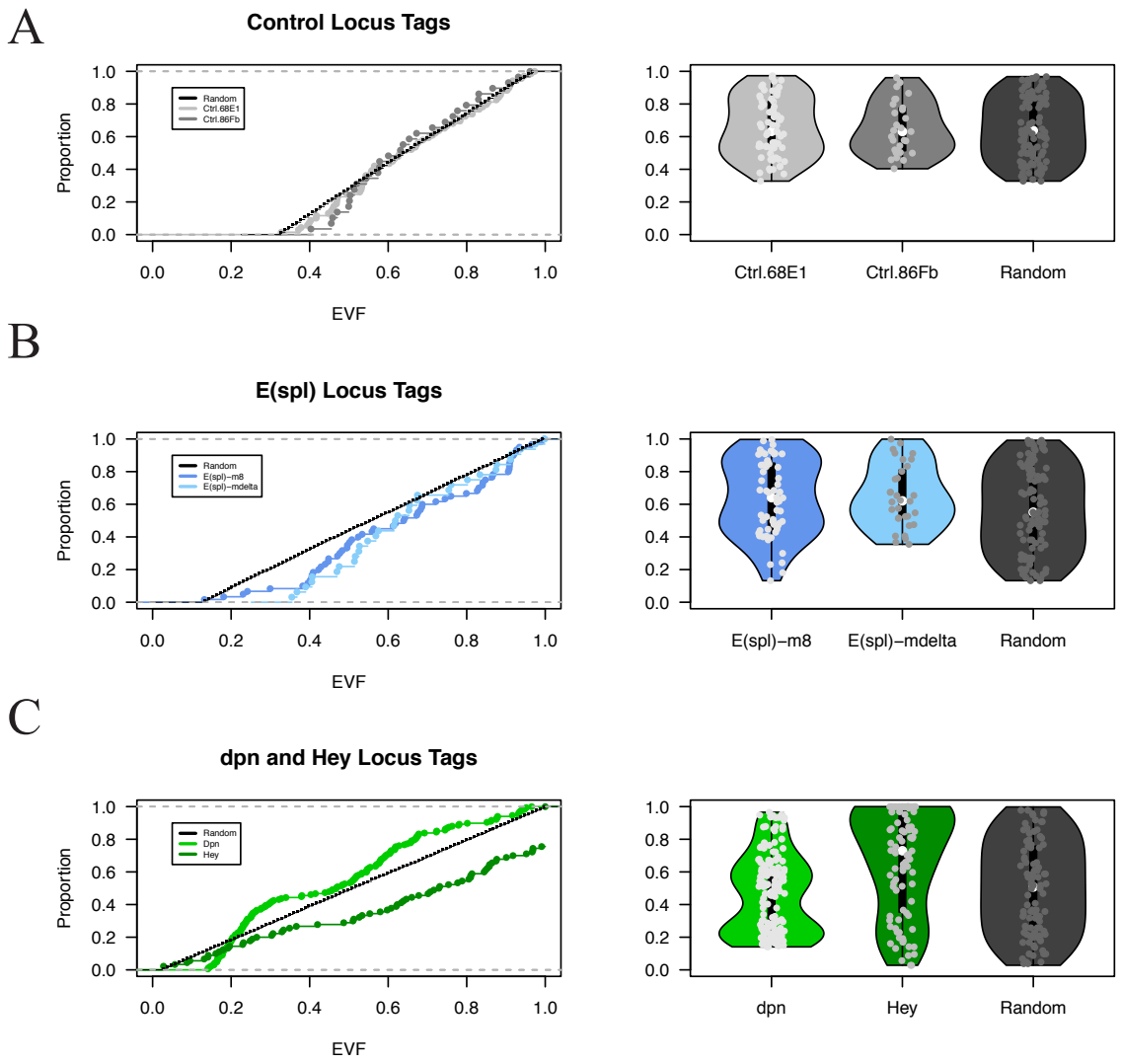


Figure 4.8: Radial position profiles for all six locus tags.

ECDF plots and violin plots showing how the relative positions for each locus tag compare to randomly generated controls with the same minimum and maximum value. The control locus tag radial positions are in (A), *E(spl)* locus tags in (B) and *dpn* and *Hey* are in (C).

The control locus tags on chromosome 3, Ctrl.68E1 and Ctrl.86Fb have a similar uniform distribution which appears to be similar to the random control (Figure 4.8A). These locus tags were not found to be significantly different from each other ($p = 0.92$) or from a random control ($p = 0.73$ for Ctrl.68E1, $p = 0.43$ for Ctrl.86Fb). In contrast, the *E(spl)* locus tags on *E(spl)-mδ* and *E(spl)-m8* were not significantly different from one another ($p=0.56$) but were significantly different from the random control (*E(spl)-mδ* $p=0.02$, *E(spl)-m8* $p = 0.01$). This effect is sensitive to the outliers in the *E(spl)-m8* dataset as when randomly sampled values were chosen excluding these, the effect ceased to be significant. The locus tags at *dpn* and *Hey* are found to be highly significantly different from each other and from the random controls (all pairwise comparisons gave a p-value for > 0.001 (see appendix 6 for table of p-values).

This statistical analysis suggests the control loci and the *E(spl)* locus tags are likely to be randomly positioned in the salivary gland nucleus, but that *dpn* and *Hey* has a significantly non-random distribution and that the two locus tags have distinct, if overlapping localisation in the salivary gland. Interestingly, *Hey*, which is on the same chromosome as *dpn* shows a reduction in point density between 0.3 and 0.5 that roughly overlaps with the region within which *dpn* gene positions are least often recorded. As these genes are linked in cis it seems natural that they might have some similarities in their profile, though the profile of *dpn* is far more pronounced.

4.2.4 Effect of ectopic Notch activation on localisation of *E(spl)-m8*.

In order to determine whether ectopic activation of Notch in the salivary gland had any effect on gene localisation the localisation of the *E(spl)-m8*.Int1 locus tag was examined under conditions where a constitutively active Notch receptor was expressed ectopically in the salivary glands via a UAS-$N^{\Delta ECD}$ transgene. Representative images of the *E(spl)-m8* locus tag in the control and UAS-$N^{\Delta ECD}$ condition are shown in Figure 4.9A. Quantification of the *E(spl)-m8* locus tag with the same EVF method as before suggests that there is an increase in the number of locus tags in the centre of the nucleus in the

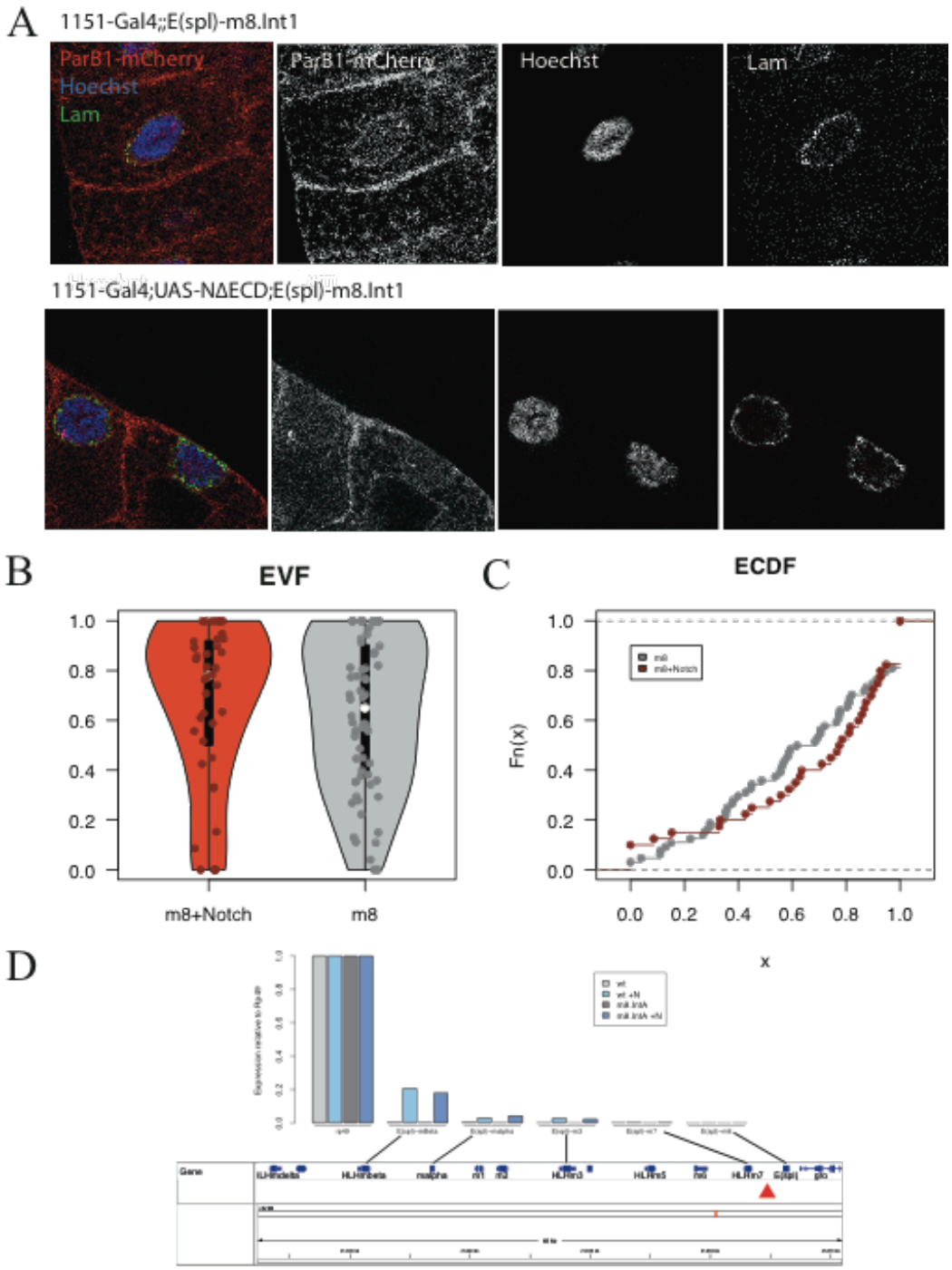


Figure 4.9: Effect of ectopic $N^{\Delta ECD}$ in the salivary gland on *E(spl)-m8* position

The *E(spl)-m8* locus tag in salivary glands with and without UAS-$N^{\Delta ECD}$ (A). Nuclear localisations of the *E(spl)-m8* locus tag (B). ECDF of locus tags with and without Notch, two sample Kolmogorov-Smirnov test p-value = 0.48 (C). qPCR measurements showing the expression of several Notch targets in response to the ectopic Notch activation in the salivary gland of wild type and locus tagged larvae with and without UAS-$N^{\Delta ECD}$ (D).

Notch activated condition (Figure 4.9B). This suggests that Notch may drive reorganisation of the genome, but more data would be required to confirm this observation as the effect was not found to be statistically significant (Kolmogorov-Smirnov test, p-value = 0.48).

In order to confirm that expression of Notch led to a change in gene expression, salivary glands from third instar larvae were dissected and the RNA levels quantified. Results confirmed that the well characterised Notch targets *E(spl)-m3*, *E(spl)-mβ* and *E(spl)-mα* all increase expression in the presence of ectopic NΔECD (Figure 4.9C). Furthermore, a similar level of response was seen when ParB1-mCherry, the locus tagging protein used for the majority of the salivary gland experiments, was present, suggesting that the locus tagging protein does not affect the ability of Notch to activate classical Notch targets at the *Enhancer of split* locus. It also indicates that although expression of other genes within the locus is highly upregulated by NΔECD, there is little effect on the genes immediately adjacent to the tag.

4.3 Discussion

The ParB-INT system was characterised for the first time in the *Drosophila* L3 larval salivary gland nuclei and it appears to produce bright and reliable labelling of individual loci. Furthermore, it was also shown that neither the genome editing procedure nor the presence of the locus tag on the chromosomes prevents the activation of classical Notch target genes at the locus tagged *E(spl)* complex suggesting that there is no major effect on the locus behaviour from these modifications.

Two possible quantification methods were devised and implemented which capture the subnuclear localisation of locus tagged genes relative to a nuclear counterstain. These were evaluated and some of the shortcomings of the first method, which dealt with absolute distances between the centroid of the nucleus and the locus tag were improved upon using a volume-based normalisation method to produce a map of relative radial distances on a continuous scale. Comparisons between the third chromosome locus tags and randomly generated control values the data suggested that the positions of the tagged third chromosome genes are largely random albeit with a preference for the nuclear interior.

Strikingly, *dpn* has two pronounced biases in its subnuclear localisation and has been imaged in the majority of cases either at the surface of the nuclear genome, facing the lamina, or on the interface between the nuclear genome and the nucleolus, both regions likely to be closely associated with lamin proteins. As mentioned earlier the gene *dpn* is known from cell culture studies to be located within a LAD (Pickersgill et al., 2006) and that Lamin A is found at the nuclear envelope and the nucleolus which makes it feasible that this positioning is linked to Lamin A or an associated factor. More recently it has been shown that in at least some cases primary sequence alone is sufficient for the organisation of chromatin into LADs (Kind and van Steensel, 2014; van Steensel and Belmont, 2017), suggesting the positioning of *dpn* could be due to a sequence motif in the vicinity of the *dpn* locus.

The finding that *E(spl)-m8* moves more centrally in the presence of ectopic Notch activation is unexpected. Nup-DamID data suggests that *E(spl)-mβ* interacts with peripheral nuclear pore proteins, (Kalverda et al., 2010) which might suggest that *E(spl)-mβ* is found peripherally, but very little was known about the actual subnuclear localisation of any Notch targets prior to this study.

The initial tests of the ParB-INT locus tagging system in this chapter prove it to be a robust and useful tool for bringing single-gene resolution to a wide range of potential imaging based approaches. The salivary gland is an interesting tissue in which to study the nuclear organisation of Notch responsive genes as in wild type animals Notch signalling activity is believed to be low or non-existent. This is supported by available expression data that report that many Notch targets have low expression in the salivary gland, that Notch is undetectable and its ligands are not expressed. Thus the majority of the data in this chapter gives a ground state picture of the subnuclear localisation of the Notch responsive genes quantified using the locus tag system.

The new locus tag insertions described in chapter 1 were confirmed to label the *Enhancer of split* locus with high specificity. The combination of negative controls and cross reactivity tests in the presence of fluorescently labelled ParB protein confirm that the locus tagging system labels *E(spl)* accurately, with a strong signal and with minimal non-specific background. Whilst the polytene nature of the chromosomes amplifies the signal by forming hundreds or even thousands of fluorescent foci on closely aligned chromosomes it also could amplify any background binding. The absence of any significant background signal suggests that the ParB-Int interaction is highly specific.

The use of a chemical DNA counterstain, Hoechst, was chosen to give nuclear measurement independent of the locus tag system. Hoechst has a high affinity for DNA and provides a good delineation of the nuclear outline with a fairly uniform staining that lends itself well to automated quantification when imaged via confocal microscopy. As shown by the images in Figure 4.9A, the locus tag signal is also preserved following an immunostaining procedure, thus widening the scope of potential applications and paving the way for studies in the central nervous system chapter described in chapter 3.

The images in Figure 4.4B are evidence that the ParB-Int system is indeed able to label two loci simultaneously in *Drosophila* nuclei as has been reported by the pioneers of the system in yeast. Interestingly, this experiment also indicated the value of the locus tagging system for potentially understanding the organisation of chromosomes at a more local level as the polytene chromosomes were observed to pair in an interdigitated manner.

All the locus tags tested on the third chromosome have very similar subnuclear localisation profiles and appear to be essentially randomly positioned. In fact, randomly simulated values confirmed that chr3 locus tags are indistinguishable from randomly positioned points within the inner 70% of the nuclear volume. The initial comparison of *E(spl)-mδ* and Ctrl-68E1 gave similar results with two different measurement methods. This suggests that these regions of the chromosome are not under a strong influence of genome organising factors, potentially indicating that these locus tags sit between LADs. This comparison also confirmed that EVF produces volume normalised measurement with less outliers, suggesting this method is sufficiently robust for the more challenging problem of detecting locus tags in neuroblast nuclei, which have a much more variable morphology and are significantly smaller.

In contrast to the third chromosome locus tags, *dpn* has a strikingly different profile which is highly statistically significant when compared to a third chromosome control. The profile is bimodal with a concentration of points towards the periphery and just peripheral of the centre. This fits with the tendency of the *dpn* locus tag to appear both at the extreme periphery of the nucleus or in very close proximity to the nucleolus. In contrast to the third chromosome locus tags this could suggest that *dpn* is located within a region that forms a LAD in the salivary gland nucleus, but also that this particular LAD can be associated with either the peripheral heterochromatin or the peri-nucleolar heterochromatin.

4.3.1 Conclusion

The key findings from this study are that all of the locus tags Ctrl.68E1, Ctrl.86Fb, *E(spl)-mδ* and *E(spl)-m8* occupy a very similar distribution of positions in the salivary gland nucleus, which is likely to be randomly determined. Secondly, the genes *dpn* and *Hey* on the second chromosome were found to be distributed in a bimodal fashion. either at the nuclear periphery or much more centrally. This distribution is quite pronounced and highly statistically significant and highlights that the subnuclear localisation of *dpn* is markedly different from the control loci and *E(spl)* complex genes. The effect was more subtle for *Hey* than for *dpn*. Given the close association of *dpn* with the lamina and the nucleolus in the imaging data in Figure 4.7 this may mean that *dpn* is located in a LAD that is stochastically positioned at either the nuclear periphery or the peri-nucleolar heterochromatin. This is supported by fact that *dpn* is present on a lamin associated gene list from a DamID study in Kc167 cells (Pickersgill et al., 2006).This may imply that there is a genetic element close to *dpn* that recruits it into LADs and peri-nucleolar heterchromatin. Finally, the *E(spl)-m8* locus tag was unexpectedly found to be localized more frequently towards the nuclear centre following the ectopic Notch activation, suggesting that the Notch signal may indeed be responsible for triggering a reorganisation of the nucleus or at least modifying the nuclear morphology in a way that the current quantification method does not account for.

5 Gene territories in neuroblast nuclei

5.1 Introduction

The neural stem cells or neuroblasts of the *Drosophila* third instar larva undergo repeated cycles of asymmetric division to produce ganglion mother cells and ultimately the neurons of the larval CNS. In order for the balance between self-renewal and differentiation to be maintained a complex network of transcription factors is used to regulate proliferation and cell fate decisions in neuroblasts and their progeny. A number of key neuroblast regulating genes are direct Notch targets, such as *Enhancer of split* complex genes and *dpn,* and loss of these together results in proliferation defects (Zacharioudaki et al., 2012). *Hey* is also a Notch target but is expressed in the neuronal progeny of neuroblasts and not the neuroblast themselves (Monastirioti et al., 2010).

Relatively little is known about the nuclear organisation of genes within the neuroblast nucleus. Although the central nervous system is densely packed with many different subtypes of neural stem cells, neurons and glial cells, methods have been developed recently that enable the isolation of specific cell types (Harzer et al., 2013). There are

also adaptations of DNA-FISH procedures specifically targeted at L3 neuroblasts as studied in this project (Larracuente and Ferree, 2015). In spite of these advances the only published study of nuclear gene position in *Drosophila* neuroblasts available at the time of writing is one that focused on the gene *hunchback* in the embryonic neuroblasts (Kohwi et al., 2013). The *hunchback* gene is known to be downregulated during the embryonic stages 10 – 13 and it was discovered that this was concomitant with a reorganisation of the nucleus that saw the *hunchback* gene repositioned from the centre to the periphery of the nucleus, the researchers concluded that this placed the locus in proximity to the heterochromatin of the nuclear lamina, facilitating and maintaining repression of transcription at the locus.

The neuroblast nuclei of the third instar CNS are an excellent system in which to study Notch targets and gene position. As mentioned above, Notch signalling components and the targets of the pathway are critical for normal neurogenesis in the fruitfly. Neuroblast nuclei are prime candidates for study of genome organisation given the lack of knowledge about gene position in this system and the relative tractability of the relatively large and numerous neuroblast nuclei. As is described later the availability of robust markers to distinguish neuroblasts from their progeny is also a great advantage. In this chapter the locus tag is tested in the larval neuroblast nucleus and the subnuclear localisations of *dpn* and *Hey*, two genes with distinct expression states and roles in the neuroblast lineage, are investigated. The locus tag assay reveals that these two genes that are found ~230kb apart on chromosome 2R (depicted in the previous chapter, in Figure 4.1 are found in different distributions along the radial axis of the neuroblast nucleus.

5.2 Results

5.2.1 The ParB-Int locus tagging system can be used to detect single copy genes in neuroblast nuclei

In order to understand the relationship between gene expression and nuclear organisation the genes *dpn* and *Hey* were chosen for a detailed analysis because whilst they are both known Notch targets in neural cell types, their expression is mutually exclusive in wild type L3 neuroblasts and neurons. Specifically *dpn* is a pan-neuroblast marker (Zacharioudaki et al., 2012) and absent from neuroblast progeny whilst *Hey* is active in a subset of neuronal progeny. Both possess a Notch sensitive enhancer although they also exhibit some Notch-independent regulation (Monastirioti et al., 2010).

Based on RNA-Seq data from FACS purified neuroblasts and neurons *dpn* is expressed over 40 fold higher than *Hey* in the neuroblast, whilst *Hey* is detected over 15 fold higher levels than *dpn* in neurons (Figure 5.1, *dpn* neuroblast FPKM = 135.5, *Hey* = 2.8, *dpn* neuron = 3.3, *Hey* = 59.6) (Harzer et al., 2013; Knoblich et al., 2012). This context allows the locus tagging system developed in the previous chapters to be used to determine whether these two genes, located within 230kb of one another on chromosome 2R, differ in their subnuclear localisation and therefore whether differential Notch responsiveness is correlated with distinct gene territories.

Before using the locus tag for analysing the gene positions, a key question was whether the genes retained their normal differential expression profiles in the presence of the locus tag. Immunostainings of brains from heterozygotes for the locus tags at *dpn* and *Hey* loci also expressing the ParB locus tagging proteins were performed to confirm that the locus tagging system does not cause aberrant expression of the tagged genes. In both cases *Dpn*+ nuclei can be seen (red) within the GFP expressing neuroblast lineages, typically surrounded by a cluster of several smaller Hey+ nuclei representing the

differentiating progeny of the neuroblasts (Figure 5.2). Thus the tag does not impede expression in their normal cell type nor does it result in derepression of these genes outside their normal cell type. This confirms that both *dpn* and *Hey* genes are expressed as usual in nuclei where the locus tagging protein is accumulated nearby to one allele of either *dpn* or *Hey*. Further, they appear to retain similar expression levels and pattern between the two conditions suggesting that the locus tagging system has little effect on the overall expression level of either gene.

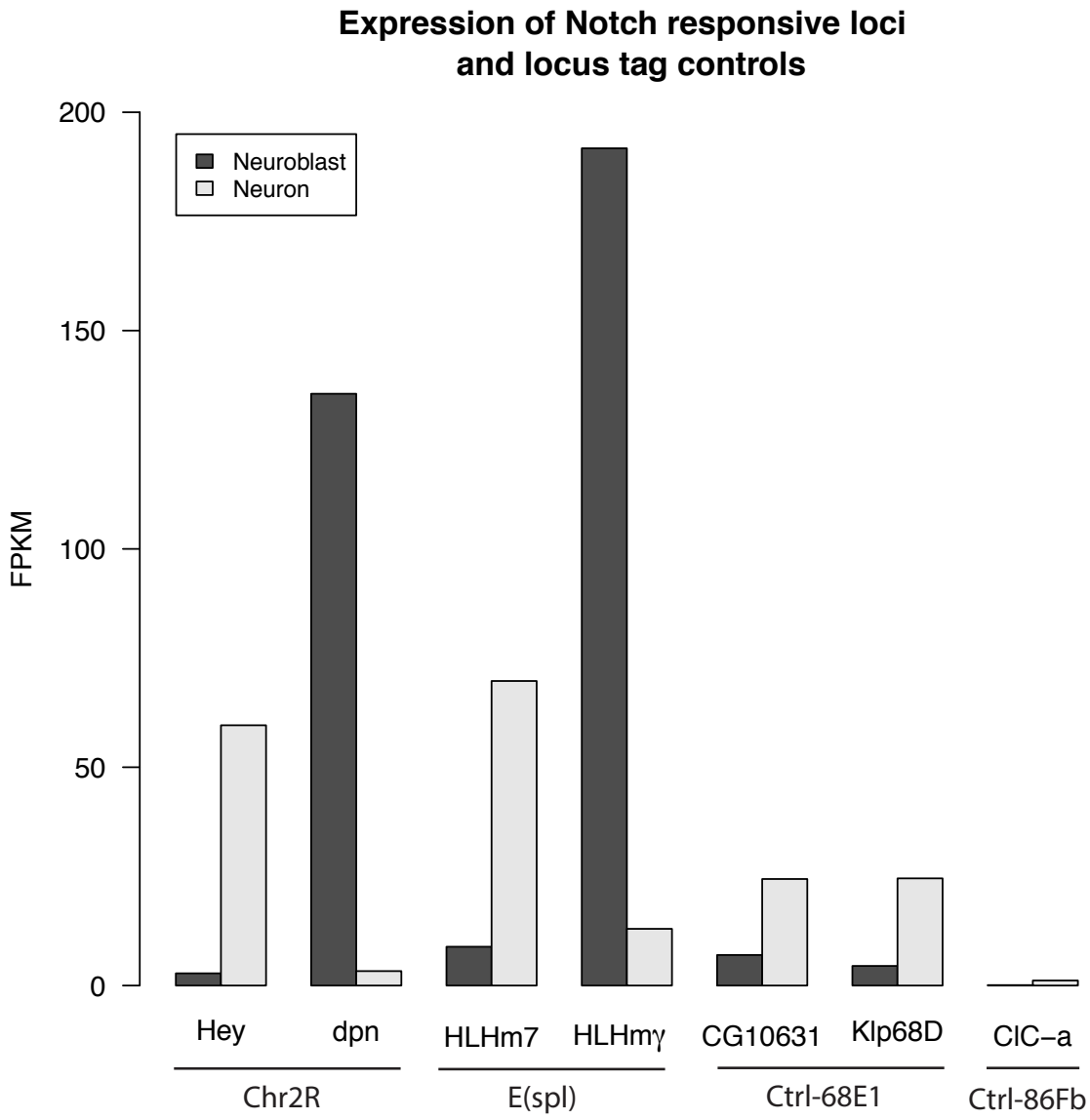


Figure 5.1: Expression levels of locus tagged genes in FACS separated CNS cell types.

Data reproduced from flybase, originally from Berger et al. 2012, was generated by performing RNA-Seq on neuroblasts and neurons from the third instar CNS following fluorescence activated cell sorting (FACS). *Hey* is expressed at low levels in the neuroblast and higher levels in the neuron, for *dpn* the inverse is true. Two *E(spl)* genes also differ markedly in their expression by cell time, whilst *E(spl)-mγ* is highly expressed in the neuroblast but *E(spl)-m7* is not, the situation is reversed in the neuron. The genes neighbouring control loci are generally of low expression level but always relatively higher in the neuron.

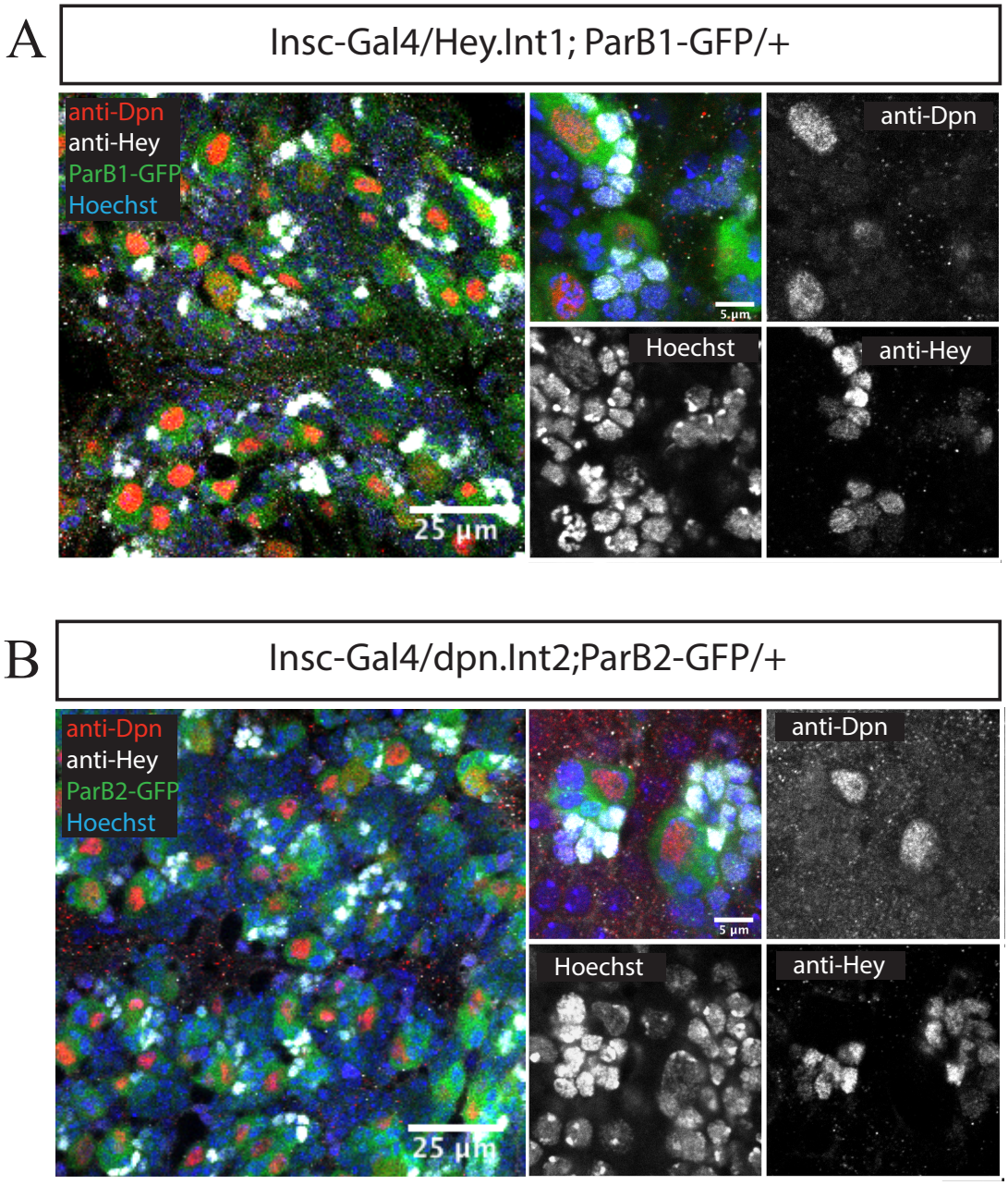


Figure 5.2: Expression of Dpn and Hey protein in neuroblasts and neurons of larvae heterozygous for the locus tag.

Immunostaining showing nuclear localisation of *Dpn* in neuroblasts (red) and Hey (white) in neurons of third instar larvae heterozygous for the *Hey*.Int1 motif insertion and expressing ParB1-GFP (green) under control of inscutable-Gal4. DNA is stained with Hoechst (blue) (A). Large image, left, digital zoom 1x, smaller image and individual channels, right, digital zoom 6x. Neuroblasts and progeny in larvae heterozygous for *dpn*.Int2 expressing ParB2-GFP under control of inscutable-Gal4 (B).

Homozygotes for transgenic locus tags also maintain typical expression of *Dpn* and Hey protein. In order to confirm normal expression of the locus tagged genes in the absence of a wild type allele, brains from individuals homozygous for the INT motif insertions were immunostained with the same anti-Hey and anti-*Dpn* antibodies as above. Imaging of the L3 neuroblast lineages in homozygotes for the *Hey*.Int1 insertion showed that Hey protein was expressed at similar levels to those observed in the neurons of a control genotype (yw;+;+) and in typical nuclear fashion (Figure 5.3A' and Figure 5.3A''). Similarly homozygotes for the *Dpn*.Int2 insertion were found to have strong nuclear staining for the *Dpn* protein as was seen in the control processed in parallel (Figure 5.3B' and Figure 5.3B''). Similar localisation and expression of Hey and Dpn in homozygotes for the locus tags indicates that the insertions of the Int motif do not cause aberrant expression of the Dpn and Hey proteins. Together with the evidence in Figure 5.2 this suggests that both the motif insertion and the accumulation of the ParB proteins at the Int motif have minimal effects of *Dpn* and Hey protein expression.

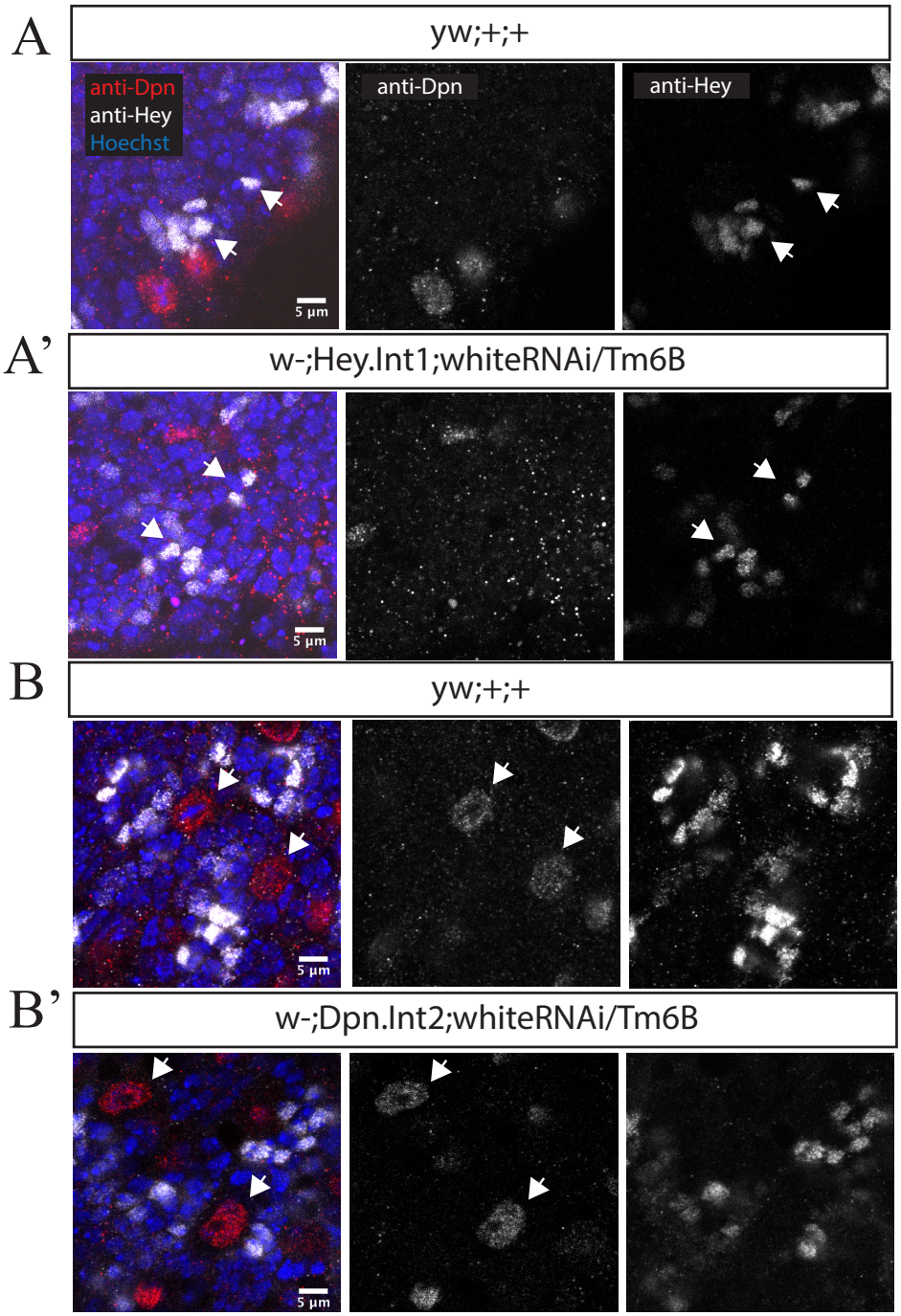


Figure 5.3: Hey and Dpn are expressed in genome-edited larvae homozygous for the locus tag insertions.

Immunostaining for Hey in brains of larvae with wild type second chromosomes (A) compared with larvae homozygous for the *Hey*.Int1 insertion (A') Hey+ neuron nuclei are visible in both genotypes with similar fluorescence intensity suggesting the locus tag insertion does not prevent expression of the gene of interest in close proximity to the tag. Control neuroblast nuclei stained with an anti-Dpn antibody (B) show similar expression of *dpn* to those homozygous for the Int2 insertion (B') suggesting the locus tag does not disrupt normal expression of *dpn*. Genotypes A': *yw;+;+,* A'' *w-;Hey.Int1;white-RNAi/+*, B'*: yw;+;+*, B'' *w-;dpn*.Int2; *white-RNAi/+*.

A second important question was whether it would be possible to detect the single copy Int insertions and whether the ParB locus tagging proteins would retain their specificity in neuroblasts, correctly recognising their cognate motifs without any cross-reaction. It was important to repeat these specificity controls as the previous chapter examined the locus tag on the polytene chromosomes and not in diploid cells. The polytene chromsomes, with their high copy number, amplified the locus tag signal due to the many identical Int motifs in close proximity to one another. Thus it was predicted that the locus tag signal would be much smaller than that seen in chapter 2.

ParB1-GFP and ParB2-GFP were expressed using an inscuteable-Gal4 line which drives expression in the neuroblast lineages. In order to confirm the locus tagging system operated with specificity and sufficient signal in the neuroblast nuclei, two ParB expressing fly stocks (Insc-Gal4; UAS-ParB1-GFP and Insc-Gal4; UAS-ParB2-GFP) were crossed to INT motif insertion lines bearing both their cognate and non-cognate motif. In this case the *Hey*.Int1 and *Dpn*.Int2 insertions were used as they are known to have different expression states in the neuroblast as explained earlier. Neuroblast nuclei were identified via expression of *Dpn* as in Figure 5.2 and Figure 5.3. Whilst the locus tag did not form band shaped fluorescent foci as seen in chapter 2, a bright round punctum was visible in each neuroblast nucleus when ParB1-GFP was paired with *Hey*.Int1 (Figure 5.4A, white arrow) and similarly when ParB2-GFP was paired with *Dpn*.Int2 (Figure 5.4B, white arrow). In contrast, no bright round puncta were observed when ParB1-GFP was paired with *Dpn*.Int2 or when ParB2-GFP was paired with *Hey*.Int1. Therefore the ParB locus tagging system characterised in the previous chapter in large, polytene nuclei also works well in neuroblast nuclei. The insertions of the Int motifs close to the *Hey* and *dpn* loci were found to produce bright, visible puncta when ParB proteins are paired with their cognate motifs but not the non-cognate partner indicating this system also functions with high specificity in the neuroblast.

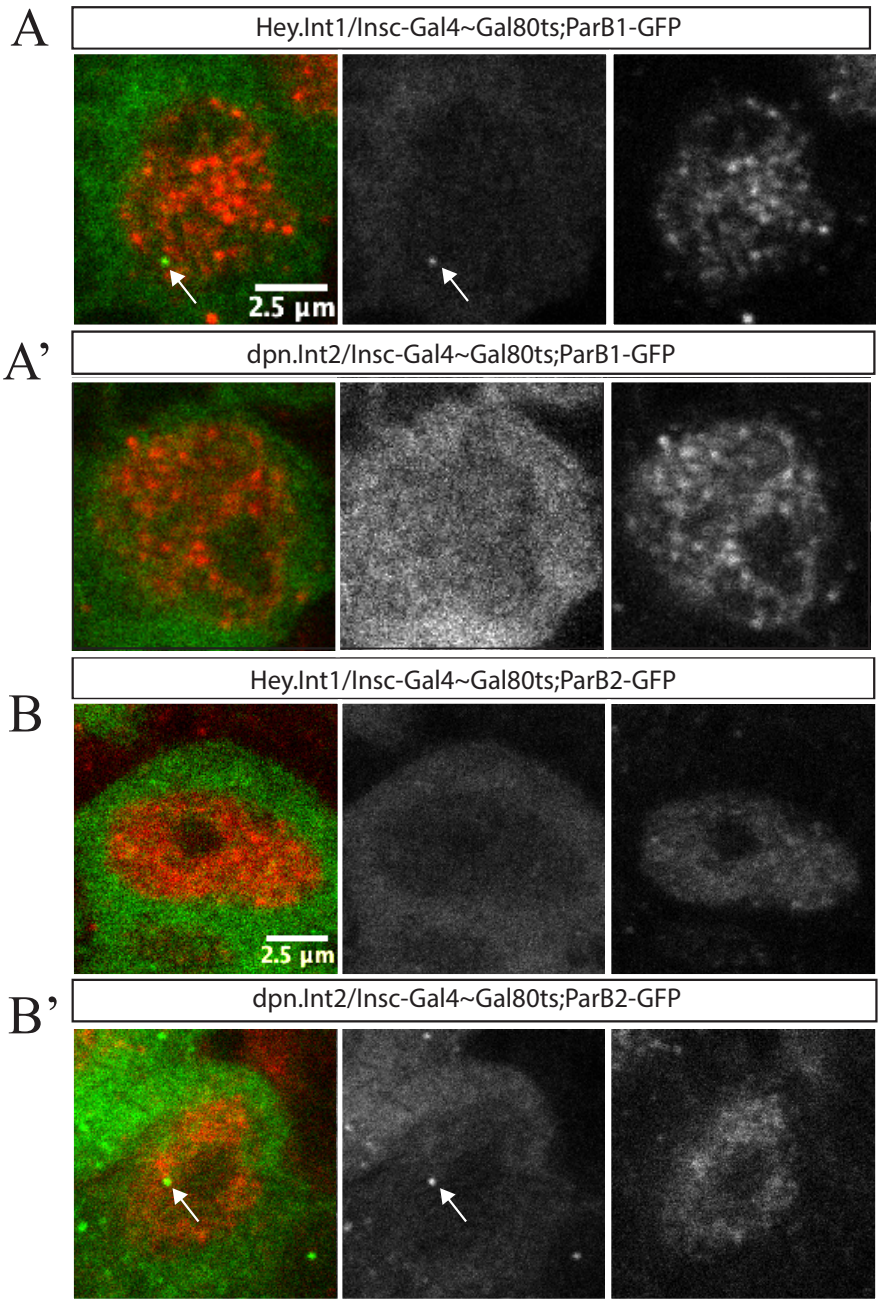


Figure 5.4: Specificity controls for the ParB-Int locus tagging system in the larval neuroblast nucleus.

Nuclei of neuroblasts expressing ParB1-GFP with the *Hey*.Int1 locus tag and the neuroblast marker *dpn* as a nuclear counterstain, the locus tag is clearly visible (white arrow) (A). Specificity control pairing ParB1-GFP with the non-cognate motif *dpn*.Int2, no locus is visible (A'). Reciprocal crosses were used to test the specificity of ParB2-GFP for the Int2 motif in the nuclei of neuroblasts. ParB2-GFP is in green no visible puncta with the *Hey*.Int1 (B). When *dpn*.Int2 is paired with the ParB2-GFP the locus tag is clearly visible (white arrow) (B').

5.2.2 Analyzing gene positions in neuroblast nuclei

In order to quantify locus tag position in a similar manner to that used in chapter 2 a marker for the nuclear interior was required. Unlike the salivary gland which is very homogenous in that it is composed of a single cell type, the neuroblasts reside in close proximity to their daughter cells these two cell types are closely packed with others such as glia. Given the specific expression of Dpn in neuroblasts, one possibility was to use the anti-Dpn antibody to specifically detect nuclear morphology of neuroblast nuclei. As can be seen in Figure 5.5 an anti-Lamin antibody and a Hoechst stain can be used to reveal all nuclei in the CNS but it is difficult to distinguish clearly the full shape of the neuroblast nuclei. The neuroblast specific anti-Dpn antibody clearly highlights the neuroblast nuclei, and it appears that the Dpn protein fills the nuclear space making detection of nuclear morphology relatively straight forward. A similar image processing approach to that in the previous chapter was used. The key adaptation is to use the fluorescence from the anti-Dpn antibody as the channel in which to detect the nuclear outline, rather than the DNA stain Hoechst which is not suitable as its staining is more diffuse in the neuroblast than in the polytene nuclei of the salivary gland. Moreover, the anti-*Dpn* stain is cell-type specific for the neuroblast, where Hoechst is not.

As the gene *Hey* is not expressed in neuroblasts, it was predicted to have a peripheral location based on previous studies of gene position. The nuclear lamina represents a repressive subdomain of the nucleus known to contain chromatin remodelling proteins that create and maintain a compacted and transcriptionally repressive chromatin state in the nuclear periphery adjacent to the lamina (Pickersgill et al., 2006; van Steensel and Belmont, 2017). In most cases where gene positions have been analysed, repressed genes have been found to be located close to the periphery. Surprisingly, the *Hey*.Int1 locus tag was detected close to the nuclear periphery in very few nuclei. The majority were found in the nuclear center (Figure 5.6A). In fact 78% of *Hey*.Int1 locus tags were recorded in the inner 50% of the nuclear volume. This bias is particularly evident in the violin plot in Figure 5.6B.

Based on most data about gene positioning, it seemed likely that the very active *dpn* gene would be positioned close to the nuclear centre. However a recent study has shown

that an ecdysone responsive gene is located near to the nuclear pore, which is thought to contribute to its regulation. In fact, the *dpn* locus fit neither of these predictions as it had a bimodal distribution. In 61% of nuclei, the *dpn.*Int2 locus tag had a position in the inner 50% of the nuclear volume, similar to the finding that the majority of *Hey.*Int1 locus tags were centrally positioned. However, in a substantial proportion of nuclei the *dpn.*Int2 puncta were observed in the periphery. This is reflected by the greater interquartile range of *dpn* locus tags on the boxplot in Figure 5.6A and there is clearly bimodal distribution when a violin plot is used as in Figure 5.6B. In order to compare the distributions, the empirical cumulative distribution frequency was calculated (Figure 5.6C). This shows that the distribution of *dpn* locus tags differs markedly in the range between 0.3 and 0.5 on the EVF scale. A two sample Kolmogorov-Smirnov test was applied to these data and the p-value, representing the likelihood that these two data sets come from distinct distributions, was 0.07. This makes it very unlikely that the peripheral bias in the distribution of the *dpn* locus tag arose due to random chance. Contrary to expectations *dpn*.Int2 appears to adopt different positions, being at the periphery in a significant fraction of nuclei. As active genes are generally thought to have a more central location, the fact that *dpn.*Int2 is sometimes peripheral was unexpected.

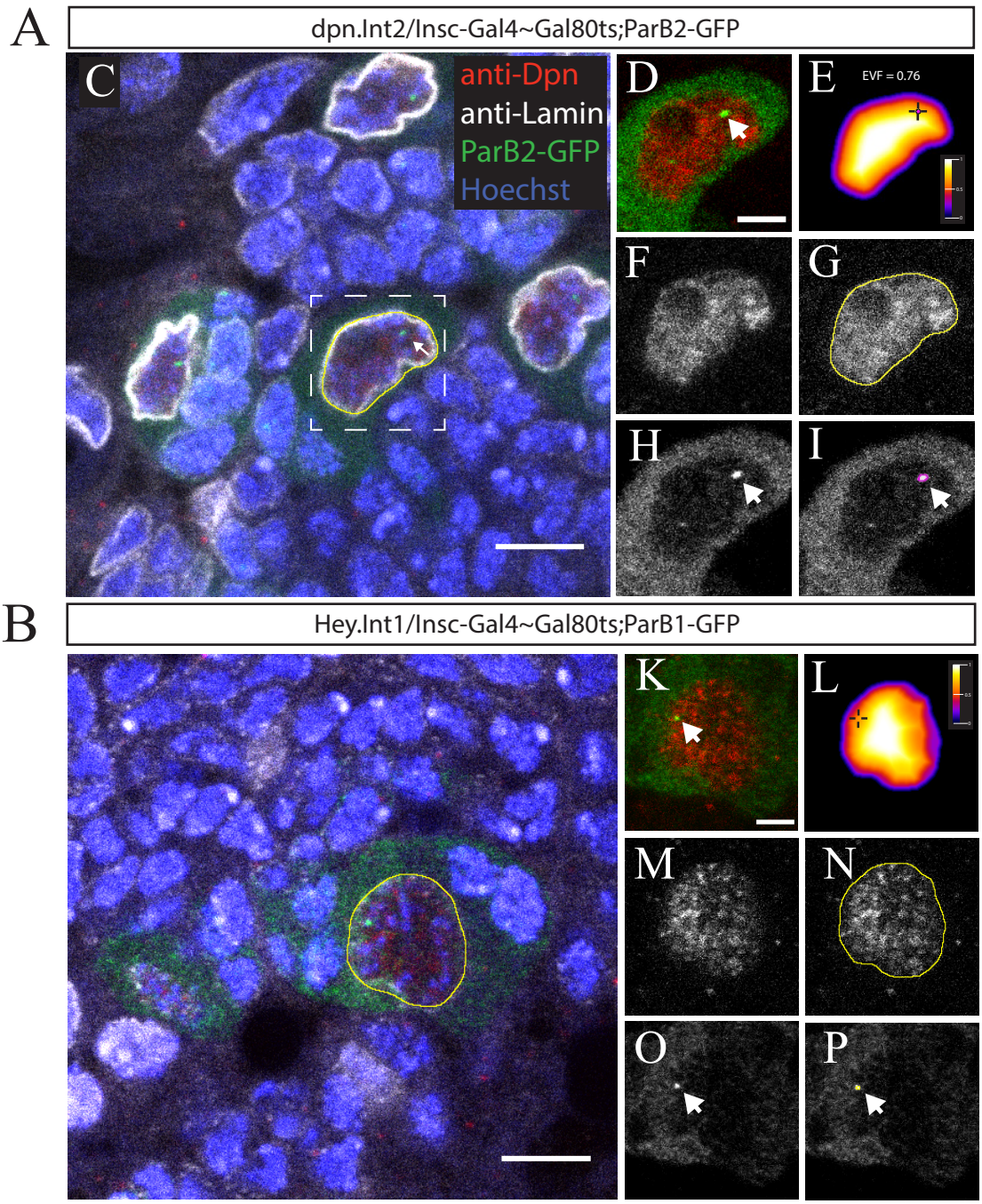


Figure 5.5: Detection of the subnuclear localisation of the *dpn* and *Hey* locus tags in neuroblasts nuclei using EVF-based distance maps.

A *dpn*+ nucleus in ParB2-GFP expressing neuroblasts from *dpn*.Int2 larvae (A). Similar representative example of ParB1-GFP expressing neuroblasts from *Hey*.Int1 larvae (B). White dashed boxes in (A) and (B) highlight zoomed in region to the right. White arrows indicate the locus tag. Locus tag and anti-Dpn merge (C) & (K), distance map of 3D segmented volume (E) & (L), scale shows 0 (periphery) to 1 (center), Black crosshair indicates XY coordinates of detected locus tag centroid. Anti-Dpn signal only (F) & (M) and with detection overlay (G) and (N) locus tag only (H) & (O) with detection overlay ((I) & (P). Scale bars 5 microns.

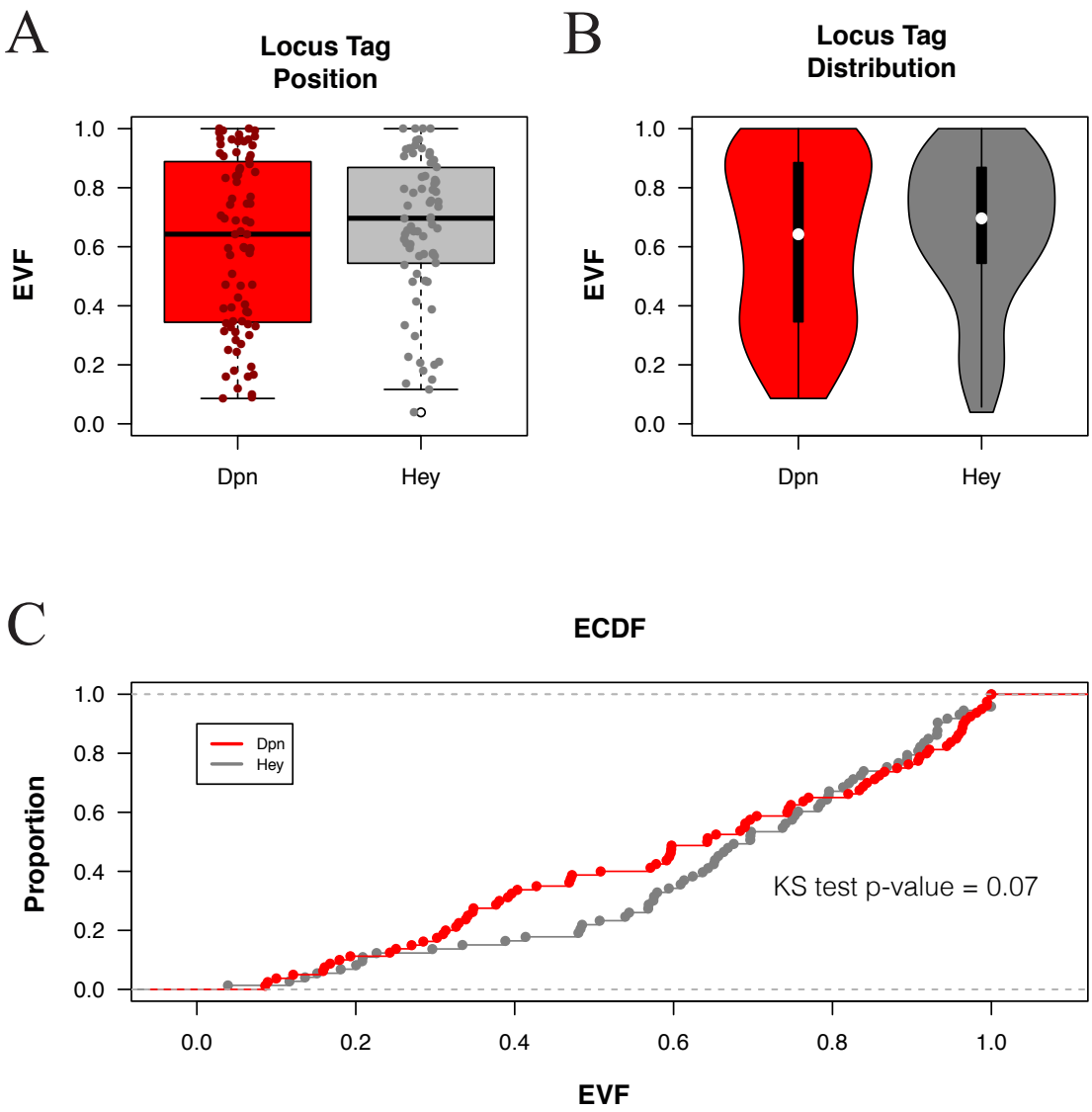


Figure 5.6: Quantification of the subnuclear localisation of the *dpn* and *Hey* locus tags in neuroblasts nuclei.

Boxplots showing the range of EVF values recorded for the *dpn*.Int2 and *Hey*.Int1 locus tags in neuroblast nuclei. Violin plots of the same values reveal that the *dpn* values have a bimodal distribution. The empirical cumulative distribution frequency curves show that the distribution of positions for the two locus tags differ markedly in the range between 0.3 and 0.6. A Kolomogorov-Smirnov test was applied and the resulting p-value was 0.07.

One possible explanation for the bimodal distribution of *dpn.*Int2 positions is that its location is affected by age or another factor, so that in some brains it would be peripheral and in others central. To assess this, the locus tag positions were colour coded according to the individual brain from which they originated. If there was a significant animal to animal variability, this would be evident from the segregation of points according to colour. The converse was the case, the colours were well mixed (Figure 5.7) indicating that the biases in the locus tag distributions do not result from locus tags within one individual brain favouring a particular position nor do any of the modal values seen on the violin plots in Figure 5.5B appear to result from one particular individual. Thus, unlike *Hey* which appears to adopt a predominantly central position in the majority of nuclei, *dpn* appears to occupy the full range of radial positions from the centre to the periphery, with a substantial peripheral bias in comparison to *Hey*.

5.2.3 Effects of Notch depletion on gene positions

The Notch pathway is active in neuroblasts and could therefore influence the position of *dpn,* which contains a Notch responsive enhancer. To investigate whether Notch activity does influence *dpn* position, an RNAi was used to deplete the Notch receptor in the neuroblasts, and the consequences for the position of *dpn* analysed. Firstly, knockdown of Notch resulted in a higher mean value for *dpn.*Int2 on the EVF scale (Figure 5.8A) suggesting *dpn* was located in the nuclear centre in more nuclei in a Notch depleted state. The bimodal distribution of the *dpn* locus tag was reproduced in the control condition (expressing white-RNAi, Figure 5.8B), but there were no longer a substantial proportion with a peripheral bias location in the Notch RNAi condition. This difference is also evident in the ECDF plot (Figure 5.8C) where the curve for the positions in the Notch RNAi nuclei lacks the peak on the left and appears shifted to the right overall. This indicates that a greater proportion of central localisations of the *dpn* locus tag in the Notch knockdown condition (KS test p-value 0.38).

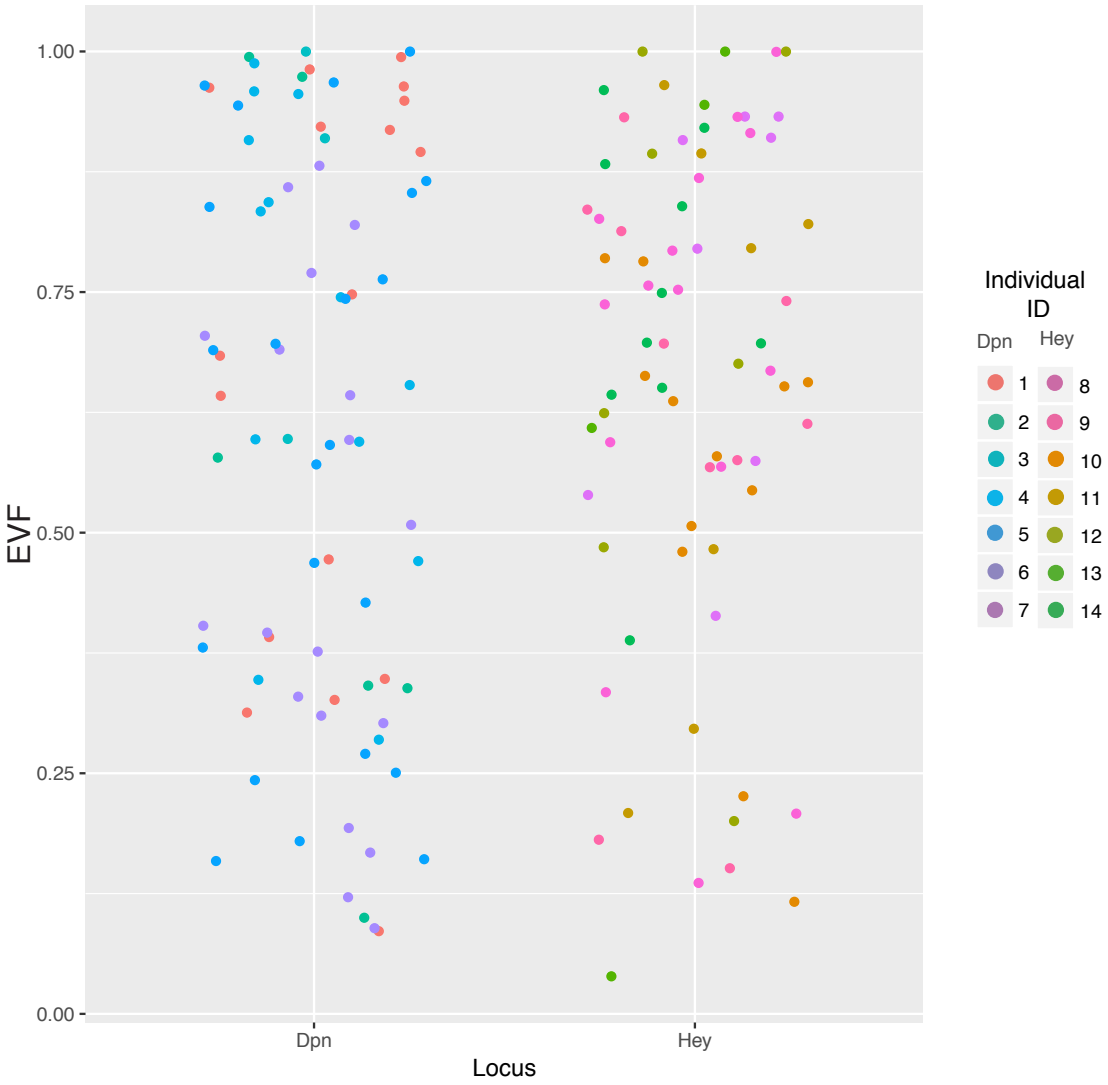


Figure 5.7: Distribution of locus tag positions in neuroblast nuclei grouped by individual.

The radial position of each locus tag imaged colour-coded by the individual brain from which it originated. Each colour is spread throughout the range of possible values indicating that the trends in gene position are unlikely to be due to variation between individuals.

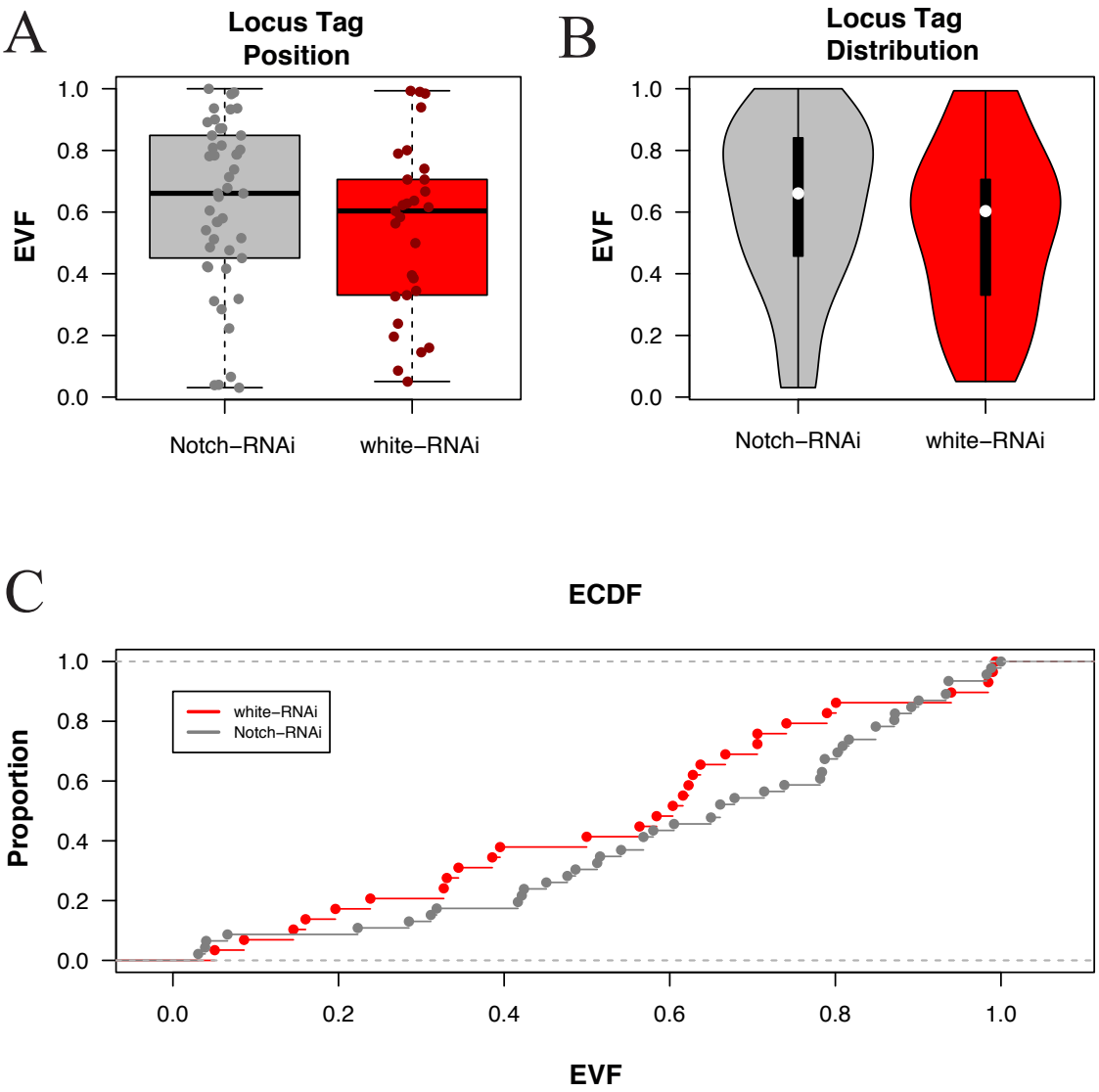


Figure 5.8: Effect of Notch knockdown on subnuclear localisation of *dpn*.Int2.

EVF values for *dpn*.Int2 locus tag in neuroblasts of Notch-RNAi (n = 46) and white-RNAi (n = 29) third instar larvae (A). Violin plot showing distribution of EVF values in (B). ECDF plot of *dpn*.Int2 with Notch-RNAi and white-RNAi (C).

The bimodal positioning of *dpn* in the neuroblast could be explained by a number of possible scenarios. Firstly each of the two alleles of *dpn* in a diploid cell could be positioned differently, and imaging larvae heterozygous for the locus tag would only reveal one position in each image but produce a bimodal distribution when the data are combined. Alternatively, the distribution could reflect a population of moving *dpn* loci, that reside longer in the periphery, and the nuclear centre than at the space in between, again resulting in a similar, bimodal distribution, but caused by each locus being dynamic, rather than two static positions.

In order to test the hypothesis that the *dpn* locus may be moving between nuclear subcompartments, neuroblast cultures were prepared from larvae expressing the locus tagging proteins and carrying Int motif insertions for *dpn* and *E(spl)-m*δ. This also served as a test of the new locus tagging tools to operate under live imaging conditions. This approach enables imaging of living neuroblasts chemically and mechanically dissociated from the CNS typically with one or two progeny cells attached. Movies were recorded overnight and many neuroblasts were observed to divide two or three times indicating healthy cultures. 3D and automated tracking was not possible due to the lack of a nuclear filling marker such as the anti-Dpn antibody used for the fixed tissue studies in Figures 5.4 – 5.8, so the *dpn*.Int2 locus tag was tracked manually on maximum projection of one movie.

Observing the *dpn.*Int2 locus over time shows that the radial position is highly variable. The locus tag is initially visible in a peripheral location at 0 minutes (Figure 5.9), but by 35 minutes the locus tag is clearly more centrally located. After a cell division at 80.5 minutes the locus tag is in the periphery once more. From here the locus tag moves slightly outwards before returning to the periphery and is positioned close to the center at 161 minutes. Prior to the second cell division, *dpn*.Int2 is closely associated with the nuclear periphery. Taken together, this behaviour indicates that the subnuclear localisation of *dpn* is highly dynamic with frequent translocations from the periphery to the nuclear centre and back in just 5-10 minutes.

The radial position of *dpn* was quantified by tracing the nuclear outline manually and generating a 2D distance map using the same principle as for the EVF approach except

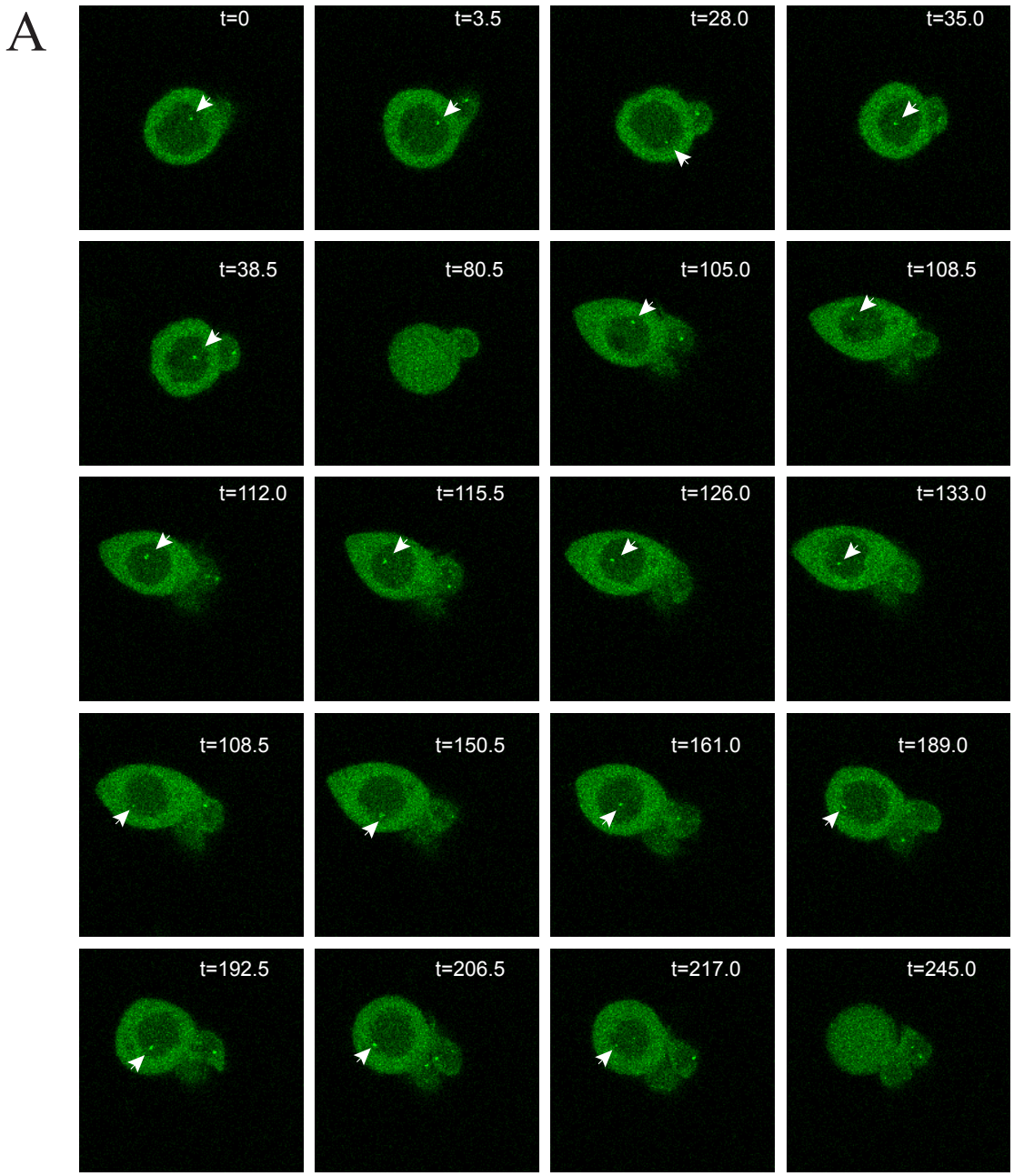


Figure 5.9: The subnuclear localisation of *dpn* is highly dynamic.

Selected frames from a movie of cultured neuroblasts dissociated from brains of *dpn*.Int2 larvae expressing ParB2-GFP. A maximum intensity projection was created of 5 z-slices was generated for each frame, two slices above and below the position of the brightest locus tag signal in the first frame. Time in minutes is indicated in white text where t=0 is the start of the movie approximately 45minutes from dissociation of the cells. White arrowheads indicate the locus tag in frames where it is visible. The neuroblast is dividing in at t=80.5 and t = 245.0. Neuroblast cultures were prepared by Dr Eva Zacharioudaki. Quantification of locus tag position in Figure 5.10

transforming a 2D area into a map of distances (Figure 5.10A). As before the scale runs from 0 in the periphery (blue) to 1 in the centre (yellow/white). The position of *dpn* was quantified over the 7.5 hour movie in every frame where the locus tag was visible (Figure 5.10B) . One notable feature of the radial position data in Figure 5.10B is that in each division cycle observed the location of the *dpn.*Int2 locus tag makes at least one journey from periphery to centre in approximately 3.5 minutes (black arrows, dashed lines indicate intervals between image acquisition). This confirms that *dpn* localisation is highly dynamic which may explain the bimodal distribution of positions observed for *dpn* in the 3D quantification presented in Figure 5.6.

Aggregating all the positional information from this movie recapitulates the population level distributions seen in the fixed tissue data. The distribution of *dpn* positions over time is bimodal, with peaks at 0.1-0.2 and 0.5-0.6 (Figure 5.10D) , whilst this differs from the exact peaks in the fixed tissue data (approximately 0.25-0.35 and 0.85-0.95 in Figure 5.6B), the distance between the modal positions is close to 0.55 in both. This agreement supports the notion that the population level distribution of the *dpn.*Int2 locus tag represents a dynamic *dpn* locus in each nucleus.

The *E(spl)-m*δ.Int1 and *dpn.*Int2 were combined to further understand the dynamics of these two loci and the test ability of the ParB-Int locus tagging system to label two loci simultaneously. Cells containing two tagged loci were imaged for 10 minutes at 1 minute intervals, the movie is shown as projections of 5 z-slices from each frame Figure 5.11A. A white arrowhead highlights the *dpn* locus tag in green and a white arrow indicates the *E(spl)-m*δ locus, whilst a white triangle denotes a daughter cell. There are in fact two adjacent puncta visible in the GFP channel (white arrow) possibly indicating that *dpn* has recently undergone replication at the beginning of this movie. During the movie it was observed that *dpn*.Int2 moved very little, but that *E(spl)-m*δ appeared to move much more rapidly within a confined region at the periphery of the nucleus close to the position of the daughter cell.

The locus tags were tracked manually to reveal the movements of the two loci over time. The tracks are shown as overlays on an averaged projection of all 11 frames (Figure 5.11B). The track for *dpn* (yellow) is relatively confined whilst tracking *E(spl)-*

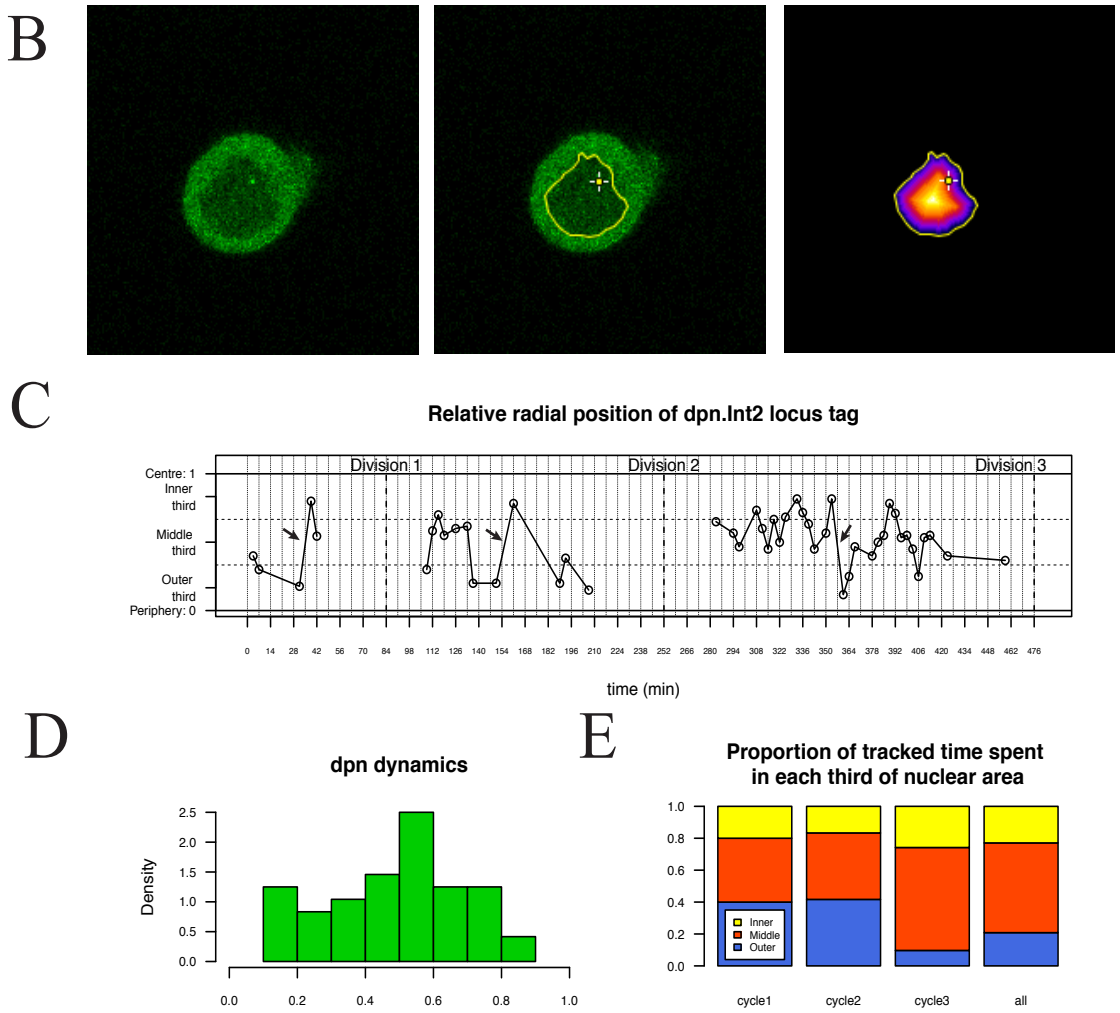


Figure 5.10: The radial position of *dpn*.Int2 oscillates in between cell divisions.

Raw image without (left) and with overlay of manually drawn nuclear outline (yellow) and locus tag selection (crosshair) (centre). 2D Distance map (right) based on nuclear outline, colour scale shows shells of equal area from periphery (blue) to centre (yellow) (A). In between divisions the radial position of *dpn*.Int2 varies from the periphery to the centre and back (B). The *dpn*.Int2 locus tag has a bimodal distribution over time, with a bias for more peripheral (0.1-0.2) and intermediate (0.5-0.6) positions (C). Proportion of time the *dpn*.Int2 locus tag was observed in the inner, middle and outer third of the nucleus (D). Neuroblast cultures were prepared by Dr Eva Zacharioudaki.

*m*δ (cyan) reveals a much more open path and a broader area is occupied in the averaged images. Whilst this could be explained by the fact that the *E(spl)-m*δ locus tag has a larger punctum, the distances plotted in Figure 5.11C were calculated from a single pixel in the centre of the locus tag and should reflect the movement of the object irrespective of the locus tag size. Both locus tags are seen to move back and forth as shown by the positive and negative gradients of the lines in Figure 5.11D but *E(spl)-m*δ covers a greater distance as judged by the height of the line and the cumulative distance plotted in Figure 5.11E.

In summary, the experiments in this chapter have characterised the ParB-Int locus tagging system in the neuroblasts of the L3 central nervous system and quantified the gene position of several Notch targets across populations and over time. The locus tagging system was found to function well in diploid cells using a standard neural lineage Gal4-driver, inscuteable-Gal4. The GFP tagged ParB proteins were used to determine the subnuclear localisation of *dpn* and *Hey* under normal Notch signalling levels, revealing that both are biased to the centre of the nucleus but that *dpn* has a peripheral bias that *Hey* does not. Furthermore, the peripheral bias of *dpn* is abolished by the use of a Notch-RNAi construct to deplete Notch in the neuroblast, converting the distribution of *dpn*.Int2 locus tag positions to a monomodal one similar to that shown by *Hey*.Int1 under normal levels of Notch. ParB.Int system was shown to label loci clearly in live cultured neuroblasts and this revealed that *dpn* is highly dynamic when followed over several hours, translocating from the nuclear centre to the periphery and back multiple times. When followed over a shorter time period, *dpn* was found to be relatively stationary and moved much less than the tagged *E(spl)-m*δ locus which moved 12 microns in 10 minutes, whilst *dpn* moved under 3 microns in the same time.

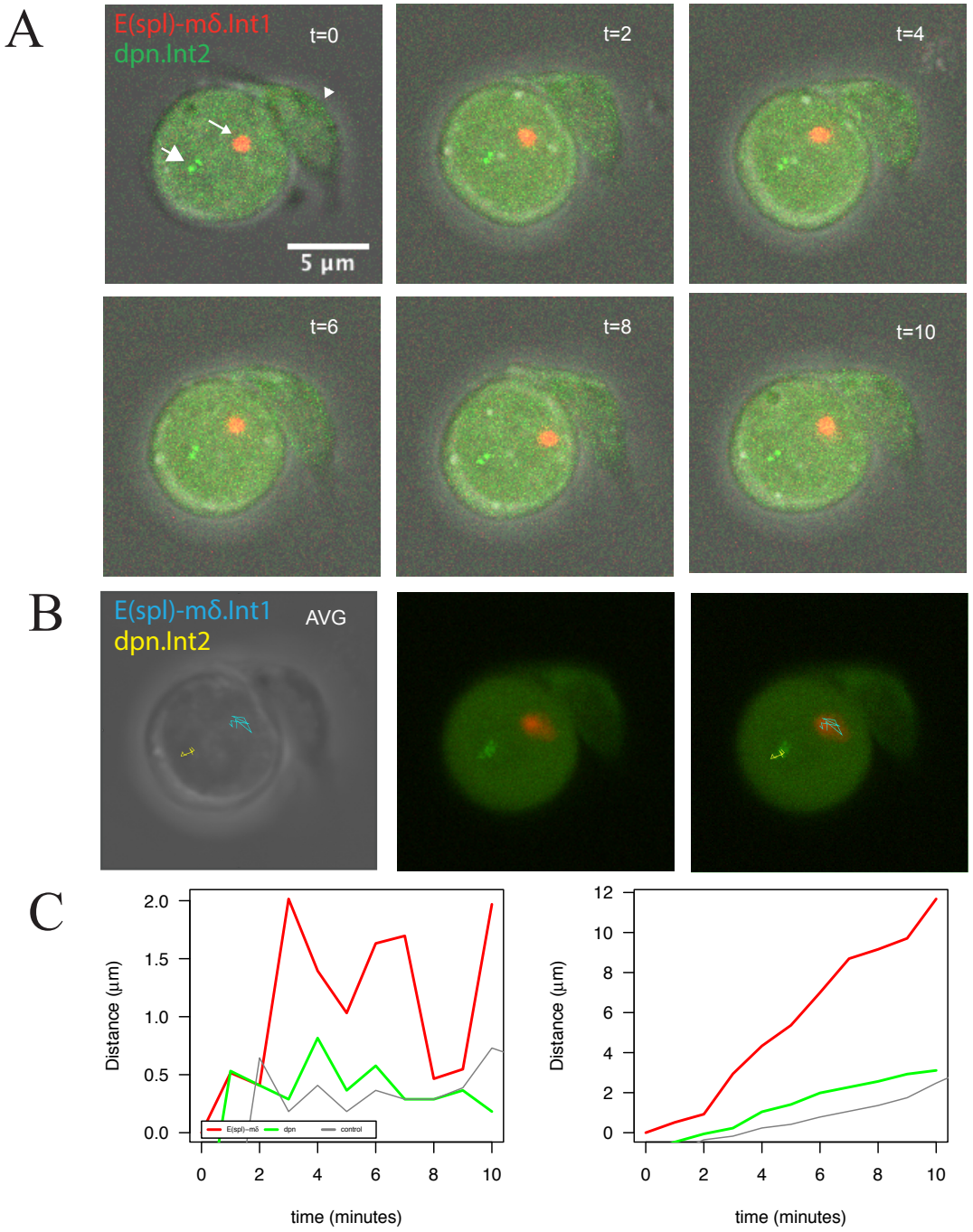


Figure 5.11: Dual labelling of *E(spl)-m*δ and *dpn* reveals distinct localisation and dynamics.

Two colour labelling of *E(spl)-m*δ (small white arrow) and *dpn* (large arrowhead) via expression of ParB1-mCherry and ParB2-GFP in cultured neuroblasts with *E(spl)-m*δ.Int1 and *dpn*.Int2 insertions (A). White triangle indicates progeny cell. Averaged projections of ten 1-minute intervals with manual tracking of locus tags overlaid. Brightfield image with tracks of *E(spl)-m*δ (cyan) and *dpn* (yellow) (left), merge of mCherry and GFP channels (centre), merge with tracks (right). Distance moved in each interval (left) and cumulative distance (right) indicate that *E(spl)-m*δ moves faster and further than *dpn* in 10 minutes. Grey line indicates tracking of the cell cell interface to control for inherent movement of the cell. Neuroblast cultures prepared by Dr Eva Zacharioudaki.

5.3 Discussion

deadpan and *Hey* have strikingly different expression states in neuroblasts, despite both having Notch responsive enhancers (Monastirioti et al., 2010; San-Juán and Baonza, 2011).Currently it is not known what accounts for their different response profiles. One possibility was that their nuclear position could explain the differential expression of these two Notch targets, which are located relatively close to each other in the genome. As a first step towards investigating this possibility, the positions of these two loci in the neuroblast nuclei were measured, using the locus tag method.

The finding that Hey has a strong preference for the centre of the nucleus was unexpected, as it is believed to be inactive in this cell type. In manually validating the 3D image segmentation it was observed a number of times that locus tags were in close proximity to a roughly circular region from which the anti-Dpn antibody was largely excluded (e.g. Figure 5.4B). It is likely that this structure is the nucleolus, which is also known to be surrounded by heterochromatin. If the central location reflects a nucleolar association then the paradox may partially be resolved, as *Hey* may therefore be associated with nucleolar heterochromatin. To investigate this, future experiments could use another marker to verify that the anti-Dpn excluded region is indeed the nucleolus. In addition, the analytic approaches would need to be adapted to monitor and how often Hey is located there. Finally, the use of antibodies to detect heterochromatin-type histone modfications (e.g. nucleophosmin and trimethylated lysine 9 on histone 3 (H3K9me3) (Dillinger and Németh, 2016)) would be important to confirm that this is a heterochromatin containing region.

As in the salivary glands, *dpn* showed a more varied distribution having a central location in some nuclei and a peripheral one in others. This suggests either a binary decision in localisation or dynamic movement of the gene, which may occur depending on the stage of the cell-cycle. Another possibility, given the neuroblast nuclei are diploid, is that there is allele specific behaviour, so that one copy of the *dpn* locus tags is concentrated in the centre of the nucleus whilst the other is located concentrated peripherally. This could be investigated by examining the distribution in nuclei with two copies of *dpn*.Int .

Given *dpn* is known to be active, and receiving a signal, Notch, from the cytoplasm, it is feasible that proximity to the nuclear pore, an organelle known to be important in gene activation, could be an important behaviour for *dpn* function. Thus potentially one *dpn* allele could be selected to be active and positioned peripherally close to the nuclear pore, whilst the other is less active, and away from the pore. In other words, the bimodal distribution of *dpn* observed in this study could be a result of observing only one allele of dpn at a time. Unfortunately, it was not possible to visualise two *dpn* locus tags simultaneously to follow up this hypothesis with the existing experimental design, due to the presence of the Gal4 driver on one of the second chromosomes and the locus tag on the other. However, using another Gal4 driver, it would be feasible to investigate the positions of both alleles in the same cell.

Support for this model comes from the shift in the distribution of the dpn locus in the Notch RNAi condition. The fact that fewer nuclei exhibited peripheral locations of *dpn* under these conditions suggests that the peripheral bias of *dpn* may be Notch dependent. The profile of the locus tag positions in the Notch knockdown condition is highly similar to that of in the control condition (compare Figure 5.5B and Figure 5.7B). Given *Hey* is known to be inactive in the neuroblast and Notch is the primary source of activation for *dpn* in the neuroblast, it seems possible that Notch is required for driving the peripheral localisation of *dpn* in some cases, but that in the absence of the signal *dpn* defaults to the behaviour of *Hey*, its relatively close neighbour.

Altogether the results have demonstrated that the locus tag method can be used to monitor gene positions in a diploid cells. The genome-editing modifications did not appear to significantly perturb the expression of the tagged loci of interest. Bright and specific puncta were detected in these cells. However, to analyze the locations the 3D segmentation pipeline described in chapter 2 had to be modified, because the neuroblast nuclei are more irregularly shaped than salivary gland nuclei. It was not possible to use the DNA stain to mark the nucleus, partly because of the close packing of cells, but the nuclear protein Dpn proved a useful method to demarcate the nuclear shape and volume. Due to the challenging morphological variation, the EVF approach was particularly important and it is unlikely that the differences in the distributions of genes would have

been found using absolute distance measurements. However, the requirement for non-parametric statistics, and consequently a large number of observations, meant that this study was underpowered to assess whether all the trends observed were significant with statistical tests.

Unfortunately it was not possible to make recombinants of the *Hey*.Int1 locus tag with the ParB1-GFP or *dpn*.Int2 with ParB2-GFP protein to compare their locations in the same nucleus, due to the fact that currently the ParB proteins are all inserted into attP sites on the third chromosomes. In the future, making ParB expressing lines on the second chromosome and the X chromosome will enable recombinants to be made with more locus tags and will enable the locus tags to be combined with more existing tools.

5.4 Conclusion

The ParB-Int system has been characterised and tested in the neural stem cells of the Drosophila CNS confirming that it is a valuable assay for gene position in diploid cell nuclei. The segmentation pipeline developed in the previous chapter has been employed similarly as before to record a volume normalised assay of 3D gene position which helps reveal the profile of localisations occupied by a population of *dpn* and *Hey* loci. A colour-coded breakdown of the measured positions confirms that the positional biases in the gene locations is independent of the individual brain imaged. Notch knockdown experiments suggest that the partial preference of *dpn* for peripheral localisations not shown by *Hey* could be driven by Notch as the profile of *dpn* is perturbed by a Notch-RNAi construct.

6 Discussion

6.1 Summary

The capability to label and visualise chromosomes in living organisms has existed for several decades though there are relatively few such tools available to study genome organisation with respect to particular genes such as Notch signalling targets. This project has succeeded in its aims to generate novel transgenic tools for achieving single gene resolution of four important Notch responsive loci via an imaging-based locus tagging approach. The ParB-Int locus tagging system has been characterised with the novel locus tag insertions in two model systems, the salivary gland nuclei, and the L3 neuroblast nuclei. In both cases ParB proteins were found to label the tagged genes of interest specifically, without cross reactivity and with no detectable effects on their normal expression. Complementary computational tools developed alongside these genetic tools enabled systematic quantification of gene position relative to nuclear size and shape. Together these tools were used as a gene position assay to determine the distribution of subnuclear positions in which HES family Notch targets are found in the nuclei of the salivary gland and neuroblast.

6.2 CRISPR/Cas9 genome-editing is a powerful tool for locus tagging

In chapter one a strategy was proposed for labelling Notch responsive genes with Int motifs to enable visualisation with the ParB-Int locus tagging system. At the time of writing there are just 3 stocks in the Bloomington fly stock collection with LacO motif repeat insertions, indicating that despite the long time for which locus tagging has been possible, there are relatively few tools to do this, and none specifically designed for studying Notch target genes. The only prior study to address gene position in *Drosophila* neuroblasts studied a single locus with DNA-FISH in embryonic neuroblast nuclei (Kohwi et al., 2013). Furthermore, in spite of the existence of a protocols paper documenting a method for using DNA-FISH in L3 neuroblasts several years prior to this study, this methods paper has no citations (Chaumeil et al., 2013). Thus there is clearly a great lack of genetic tools for studying genome organisation with locus tag technologies. This study provides a strategy for tagging more genes as well as four novel locus tagged loci that are important developmental genes.

6.3 The ParB-Int locus tagging system is versatile and robust.

In this study the ParB-Int locus tagging system was successfully used to image genes in both polytene and diploid cell nuclei. This was achieved with different Gal4 driver lines and combined with DNA stains and various antibodies. Despite fixation and immunostaining the native locus tag fluorescence was detectable without the need for antibodies to amplify the signal even in the diploid nuclei. The realisation of dual colour live imaging of two HES family Notch targets and the lack of detectable perturbations to expression of the tagged genes illustrates that the ParB-Int system is preferable over DNA-FISH and traditional locus tagging approaches, such as LacI-GFP/LacO, in many respects. As mentioned in the introduction, DNA-FISH is often challenging to combine with immunostainings and it is relatively more expensive, whilst genetically encoded locus tags can be used repeatedly at very little cost. The cell-type specificity afforded by

the Gal4/UAS system used to activate the ParB expression is also potentially a very useful aspect of the system, especially in tissues comprised of multiple densely packed cell types such as the third instar CNS.

6.4 HES family genes have distinct subnuclear localisations in salivary gland nuclei.

The *dpn* and *Hey* locus tags were found to occupy a range of positions that fell into a biomodal distribution when data from many nuclei was aggregated. This indicates that these genes are non-randomly positioned in the salivary gland nucleus. In contrast the distribution of *E(spl)* locus tag positions was not found to be strongly bimodal, lacking the strongly peripheral positioning observed for the *dpn* locus tag. However the *E(spl)* locus tags positions were found to be significantly different from randomly generated control positions and are thus also observed to be non-randomly positioned although to a lesser degree. One previous study into the relationship between the organisation of salivary gland chromosomes and tissue specific gene regulation concluded that the 3D organisation of the genome was unlikely to contribute to gene regulation (Hochstrasser et al., 1986). The lack of consistency between the subnuclear compartments occupied by the four HES family genes tagged in this study is therefore compatible with this conclusion. Moreover, the dual role of Su(H) as a repressor in the absence of NICD and an activator in the presence of NICD and Mam (Bray, 2006) makes it feasible for Notch targets to be in a stably repressed state if they are bound by Su(H), or alternatively their regulation is dominated by other repressive chromatin factors whose effects are amplified by the high copy number of polytene chromosomes (up to 1024 chromosome copies are aligned together (Mortin and Sedat, 1982)). Thus, in the salivary gland nuclei, the organisational requirements of fitting large chromosomes into the nucleus likely dominate the structural arrangement of the genome, whereas in diploid cells regulatory subcompartments may be more important.

The localisation of *dpn* at both the nuclear periphery and the perinucleolar region argues that there is some degree of flexibility in the organisation of lamina associated domains (LADs). This is supported by one study that argues the peripheral heterochromatin, pericentriolar heterochromatin and perinucleolar heterochromatin should all be

considered one functional compartment for late replicating heterochromatin (Ragoczy et al., 2014). Further, another study found stochastic positioning of LADs in mammalian cells and identified the Lamin A being the only Lamin isoform found at peri-nucleolar LADs, whilst Lamin A, B1 and B2 were found associated with LADs in the nuclear periphery (Kind and van Steensel, 2014; Kind et al., 2013). This suggests that stochastic positioning of LADs is a conserved feature of genome organisation between mammals and *Drosophila* and present in both diploid and polytene nuclei and that the *dpn*.Int2 locus tag may be a useful tool for studying LADs in the future.

6.5 *Hey* and *dpn* are differentially regulated and differentially positioned in the neuroblast nucleus.

Contrary to the expectation that the gene *Hey,* which is inactive in the neuroblast, would be positioned in the nuclear periphery, it was found to reside predominantly in the nuclear centre. The localisation of *Hey* differed from its active neighbour *dpn* in that the latter was found in the periphery more often than *Hey.* This suggests that localisation to a peripheral subnuclear compartment may be a key factor that distinguishes *dpn* from *Hey* and may explain why *dpn* is highly active in the neuroblast but *Hey* is not. (Knoblich et al., 2012). This idea is supported by the evidence that *dpn* translocates frequently between the nuclear periphery and the nuclear centre in cultured neuroblasts. The loss of the peripheral bias in *dpn* localisation when Notch is depleted suggests that Notch is required for localisation of *dpn* to the nuclear periphery, but live imaging of neuroblasts in a Notch-RNAi condition would be required to prove that Notch signalling is required for the active translocation of *dpn* to the periphery. The absence of any bias for peripheral localisation in the profile of *Hey* also implies that a central, repressive subcompartment may be important for maintaining *Hey* in a repressed state in the neuroblast which could be the peri-nucleolar heterochromatin, though this hypothesis was not tested in this study.

The gene *dpn* is well established as a neuroblast marker, an important regulator of neuroblast proliferation and has been shown to be directly regulated by Notch by an

enhancer containing three Su(H) binding sites (San-Juán and Baonza, 2011). Why this gene is active yet another nearby Notch target, *Hey* is inactive, is an important question in understanding Notch regulated genes (Monastirioti et al., 2010). This study provides the first evidence that differential subnuclear localisation correlates with the differential expression of the two genes in the neuroblast and also shows that *dpn* translocates to the nuclear periphery, suggesting that the differential expression of *dpn* with respect to *Hey* may be a result of differential gene dynamics.

6.6 Dual labelling of *dpn* and *E(spl)-mδ* reveals they have distinct short term dynamics.

Imaging of *dpn* and *E(spl)-mδ* simultaneously revealed that they can occupy distinct gene territories and have different dynamics over a short, 10 minute interval despite the fact that both regions are Notch responsive in this cell type. This experiment demonstrated the potential of the ParB-Int locus tagging system and the novel Int insertions generated in this study. Whilst *dpn* and *E(spl)* genes such as *E(spl)-mγ* (which is a direct neighbour of the locus tag at *E(spl)-mδ*) have been found to be required redundantly for normal proliferation of the neuroblast (Zacharioudaki et al., 2012), this movie shows the two genes occupy distinct gene territories and have different dynamics. Future studies will be able to use the tools developed in this study to determine whether *E(spl)* genes undergo translocations between the nuclear periphery and the nuclear centre as *dpn* does.

6.7 Limitations of this thesis

Whilst this thesis reports the discovery of interesting and unexpected features in the organisation of Notch responsive genes, one key limitation of the experimental design is that without further tools or markers the locus tagging system cannot distinguish between nuclear subcompartments such as the nuclear pore or the lamina. Hence, the study focused on the radial profile of the locus tagged genes to determine the gene territories that each locus occupies in the nucleus. The unexpected evidence that peri-

nucleolar heterochromatin might be important in at least the regulation of Notch target organisation and also possibly expression from that locus meant that a marker specific for the nucleolus was lacking throughout the study, which would have been valuable in following up the observation of locus tags associated with perinucleolar heterochromatin and assisted quantifying such associations.

This project aimed to make quantification of the locus tag more systematic and less biased than purely manual quantifications. The use of automatic thresholding ensures each nucleus is segmented taking into account local variations in fluorescence intensity and means that the analysis is easily reproducible where manual selection of thresholds may have been biased and difficult to extend to additional datasets. The use of the EVF method means that in future multiple cell types could be compared on the same scale, even if the nuclei have different size and shape. The pipeline was not purely automated as some background fluorescence did exist and so each locus tag detection was validated manually whilst being detected and quantified automatically. This enabled nuclei with more or less than one locus tag detection to be excluded, as well as images where staining was of poor quality or detection was incorrect (such as when two nuclei were positioned very close together and detected as one object) to be discarded.

Despite these advantages, there are also some important limitations of the image processing pipeline. Firstly, the EVF scale produces non-normally distributed data varying between 0 and 1 (a Gaussian distribution extends to + and – infinity) meaning parametric statistical tests requiring normal data such as t tests and ANOVA are not applicable. The non-parametric Kolmogorov-Smirnov test is appropriate for this application, but it does typically require a higher number of data points to achieve statistical significance than a parametric test would which requires more samples and more images. The approach developed here uses two different nuclear stains for different cell types which both have different optical properties and staining or immunoreactivity properties. Quantification could be made more accurate by imaging the nucleus with more than one marker, such as an additional lamin marker which would not only highlight the nuclear envelope but provide a second measurement of the extent of the nuclear periphery. Whilst a lamin staining was included with some images

collected for this study, it was generally too variable to use to reliably detect nuclear dimensions via the 3D segmentation approach developed in this study.

Depletion of the Notch receptor for the experiments detailed in chapter 5 was performed using an RNA interference construct. Whilst a suitable RNAi control was used to rule out non-specific RNAi mediated effects, the original characterisation of this tool shows that its effects vary depending on the Gal4 driver used (Presente et al., 2002). Thus it is not known how effective the depletion of Notch was, but the fact that a change in the subnuclear localisation of *dpn* was detected suggests that it was substantial. In the future, antibody staining for the Notch receptor or fluorescent in situ hybridisation to detect Notch mRNA levels may be useful approaches to combine with the locus tag assay to determine the knockdown efficiency on a cell by cell basis.

The vast majority of the experiments detailed in this thesis were performed on fixed tissue using formaldehyde. Whilst necessary for immunohistochemistry protocols, and widely used in ChIP and chromatin conformation capture studies it is possible that fixation may result in artefacts that confound the detection of some features of chromosomal organisation (Mortin and Sedat, 1982). Despite the number of advantages to working with fixed tissue, this prevents the possibility to record time series of images and to determine the dynamics of locus movement and potential changes in gene position throughout the cell cycle. In this study, live imaging has revealed that at least some Notch responsive genes are highly dynamic, arguing that future efforts focused on using these tools in live imaging studies and developing the range of genetic tools for live imaging will likely aid our understanding of nuclear organisation and gene regulation greatly.

6.8 Future Directions

The success of the ParB-Int system and relative ease with which it can be retargeted using CRISPR/Cas9 means that tagging more loci with Int motifs would greatly expand the number of biological questions that could be asked with a relatively simple assay. To further understand the relationship between the nuclear pore and the Notch pathway,

such efforts could focus on Notch responsive loci found to interact with nuclear pore complex proteins in previous studies. The locus tags created in this study enable the study of chromatin composition at *E(spl)* and changes following Notch activation with a wide range of approaches from immunostaining to live imaging. The 3D segmentation pipeline developed in this study could be adapted into a screen for factors driving nuclear organisation or to quantify changes before and after Notch signalling events with relative ease. Live imaging of Notch targets in different tissues would be a fascinating insight into how genome organisation and dynamics vary during differentiation and between a range of cell types with and without active Notch signalling. Finally, the combination of chromosomal locus tags with live transcript tags such as the MS2 system could potentially reveal the relationship between nuclear organisation and gene expression in a real-time and quantifiable manner(Arib et al., 2015; Gallardo and Chartrand, 2011).

6.9 Conclusion

This study established a CRISPR/Cas9 genome-editing procedure and combined it with ΦC31 integrase transgenesis to tag important Notch responsive genes with a novel locus tagging system. Following validation of the genome editing and Int insertions, specificity of the locus tag in the salivary gland nucleus was confirmed via control crosses. Distinct nuclear organisation profiles were detected for different Notch targets using a purpose-built 3D image segmentation pipeline. This method was used to show that whilst control locus tags and those at *E(spl)* loci appear to be positioned randomly in the salivary gland nuclei, *dpn* and to a lesser extent *Hey* are found to be positioned in a non-random manner. In the neuroblast, the differential activity levels of *Hey* and *dpn* were found to correlate differential subnuclear localisation. Specifically a peripheral bias shown in the distribution of *dpn* localisations that was not observed for *Hey,* and was abolished by Notch depletion. Simultaneous live imaging of the *dpn* and *E(spl)-mδ* locus tags revealed that these also genes differ in their localisation and dynamics.

7 References

Abruzzi, K. C., Belostotsky, D. A., Chekanova, J. A., Dower, K. and Rosbash, M. (2006). 3'-end formation signals modulate the association of genes with the nuclear periphery as well as mRNP dot formation. *EMBO J.* **25**, 4253–4262.

Ahmed, S., Brickner, D. G., Light, W. H., Cajigas, I., McDonough, M., Froyshteter, A. B., Volpe, T. and Brickner, J. H. (2010). DNA zip codes control an ancient mechanism for gene targeting to the nuclear periphery. *Nature Cell Biology* **12**, 111–118.

Akhtar, A. and Gasser, S. M. (2007). The nuclear envelope and transcriptional control. *Nat Rev Genet* **8**, 507–517.

Arib, G., Cléard, F., Maeda, R. K. and Karch, F. (2015). Mechanisms of Development. *Mech. Dev.* **138**, 133–140.

Babaoğlan, A. B., Housden, B. E., Furriols, M. and Bray, S. J. (2013). Deadpan Contributes to the Robustness of the Notch Response. *PLoS ONE* **8**, e75632.

Ballester, M., Kress, C., Hue-Beauvais, C., Kiêu, K., Lehmann, G., Adenot, P. and Devinoy, E. (2008). The nuclear localization of WAP and CSN genes is modified by lactogenic hormones in HC11 cells. *J. Cell. Biochem.* **105**, 262–270.

Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A. and Cavalli, G. (2011). Polycomb-Dependent Regulatory Contacts between Distant Hox Loci in Drosophila. *Cell* **144**, 214–226.

Belmont, A. S. and Straight, A. F. (1998). In vivo visualization of chromosomes using lac operator-repressor binding. *Trends in Cell Biology* **8**, 121–124.

Berger, A. B., Cabal, G. G., Fabre, E., Duong, T., Buc, H., Nehrbass, U., Olivo-Marin, J.-C., Gadal, O. and Zimmer, C. (2008). High-resolution statistical mapping reveals gene territories in live yeast. *Nat Meth* **5**, 1031–1037.

Bischof, J., Maeda, R. K., Hediger, M., Karch, F. and Basler, K. (2007). An

optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 3312–3317.

Blobel, G. (1985). Gene gating: a hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 8527–8529.

Boettiger, A. N., Bintu, B., Moffitt, J. R., Wang, S., Beliveau, B. J., Fudenberg, G., Imakaev, M., Mirny, L. A., Wu, C.-T. and Zhuang, X. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422.

Boyle, S., Rodesch, M. J., Halvensleben, H. A., Jeddeloh, J. A. and Bickmore, W. A. (2011). Fluorescence in situ hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis. *Chromosome Res.* **19**, 901–909.

Bray, S. (1998). Notch signalling in Drosophila: three ways to use a pathway. *Semin. Cell Dev. Biol.* **9**, 591–597.

Bray, S. J. (2006). Notch signalling: a simple pathway becomes complex. *Nat. Rev. Mol. Cell Biol.* **7**, 678–689.

Bray, S. J. (2016). Notch signalling in context. *Nat. Rev. Mol. Cell Biol.* **17**, 722–735.

Brickner, D. G., Ahmed, S., Meldi, L., Thompson, A., Light, W., Young, M., Hickman, T. L., Chu, F., Fabre, E. and Brickner, J. H. (2012). Transcription Factor Binding to a DNA Zip Code Controls Interchromosomal Clustering at the Nuclear Periphery. *Developmental Cell* **22**, 1234–1246.

Capelson, M., Liang, Y., Schulte, R., Mair, W., Wagner, U. and Hetzer, M. W. (2010). Chromatin-Bound Nuclear Pore Components Regulate Gene Expression in Higher Eukaryotes. *Cell* **140**, 372–383.

Chaumeil, J., Micsinai, M. and Skok, J. A. (2013). Combined Immunofluorescence and DNA FISH on 3D-preserved Interphase Nuclei to Study Changes in 3D Nuclear Organization. *JoVE*.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*.

Cremer, T. and Cremer, C. (2006). Rise, fall and resurrection of chromosome territories: a historical perspective. Part II. Fall and resurrection of chromosome territories during the 1950s to 1980s. Part III. Chromosome territories and the functional nuclear architecture: experiments and models from the 1990s to the present. *Eur J Histochem* **50**, 223–272.

Dang, H., Lin, A. L., Zhang, B., Zhang, H.-M., Katz, M. S. and Yeh, C.-K. (2009). Role for notch signaling in salivary acinar cell growth and differentiation. *Dev. Dyn.* **238**, 724–731.

de Wit, E. and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* **26**, 11–24.

Dekker, J. and Mirny, L. (2016). Perspective. *Cell* **164**, 1110–1121.

Delidakis, C., Monastirioti, M. and Magadi, S. S. (2014). *E(spl): Genetic, Developmental, and Evolutionary Aspects of a Group of Invertebrate Hes Proteins with Close Ties to Notch Signaling*. 1st ed. Elsevier Inc.

Dillinger, S. and Németh, A. (2016). Quantitative Immunofluorescence Analysis of Nucleolus-Associated Chromatin. *Methods Mol Biol* **1455**, 59–69.

Djiane, A., Zaessinger, S., Babaoğlan, A. B. and Bray, S. J. (2014). Notch Inhibits Yorkie Activity in Drosophila Wing Discs. *PLoS ONE* **9**, e106211.

Doe, C. Q. (2008). Neural stem cells: balancing self-renewal with differentiation. *Development* **135**, 1575–1587.

Dubarry, N., Pasta, F. and Lane, D. (2006). ParABS Systems of the Four Replicons of Burkholderia cenocepacia: New Chromosome Centromeres Confer Partition Specificity. *Journal of Bacteriology* **188**, 1489–1496.

E C Lai, R. B. A. J. W. P. (2000). The Enhancer of split Complex of Drosophila includes four Notch-regulated members of the Bearded gene family. 1–15.

Eagen, K. P., Hartl, T. A. and Kornberg, R. D. (2015). Stable Chromosome Condensation Revealed by Chromosome Conformation Capture. *Cell* **163**, 934–946.

Efstratiadis, A., Szabolcs, M. and Klinakis, A. (2014). Notch, Myc and Breast Cancer. *Cell Cycle* **6**, 418–429.

Egger, B., Chell, J. M. and Brand, A. H. (2008). Insights into neural stem cell biology from flies. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 39–56.

Fallah, Y., Brundage, J., Allegakoen, P. and Shajahan-Haq, A. N. (2017). MYC-Driven Pathways in Breast Cancer Subtypes. *Biomolecules* **7**, 53.

Filion, G. J., van Bemmel, J. G., Braunschweig, U., Talhout, W., Kind, J., Ward, L. D., Brugman, W., de Castro, I. J., Kerkhoven, R. M., Bussemaker, H. J., et al. (2010). Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. *Cell* **143**, 212–224.

Furman, D. P. and Bukharina, T. A. (2008). How Drosophila melanogaster Forms its Mechanoreceptors. *Curr. Genomics* **9**, 312–323.

Gadal, O. (2016). jcs.188250.full. 1–55.

Gall, J. G. (2016). The origin of in situ hybridization – A personal history. *Methods* **98**, 4–9.

Gallardo, F. and Chartrand, P. (2011). Visualizing mRNAs in fixed and living yeast cells. *Methods Mol Biol* **714**, 203–219.

Gibcus, J. H. and Dekker, J. (2013). The Hierarchy of the 3D Genome. *Molecular Cell* **49**, 773–782.

Gratz, S. J., Ukken, F. P., Rubinstein, C. D., Thiede, G., Donohue, L. K., Cummings, A. M. and O'Connor-Giles, K. M. (2014a). Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. *Genetics* **196**, 961–971.

Gratz, S. J., Wildonger, J., Harrison, M. M. and O'Connor-Giles, K. M. (2014b). CRISPR/Cas9-mediated genome engineering and the promise of designer flies on demand. *fly* **7**, 249–255.

Hartenstein, A. Y., Rugendorff, A., Tepass, U. and Hartenstein, V. (1992). The function of the neurogenic genes during epithelial development in the Drosophila embryo. *Development* **116**, 1203–1220.

Harzer, H., Berger, C., Conder, R., Schmauss, G. and Knoblich, J. A. (2013). FACS purification of Drosophila larval neuroblastsfor next-generation sequencing. *Nat Protoc* **8**, 1088–1099.

Hochstrasser, M., Mathog, D., Gruenbaum, Y., Saumweber, H. and Sedat, J. W. (1986). Spatial organization of chromosomes in the salivary gland nuclei of Drosophila melanogaster. *The Journal of Cell Biology* **102**, 112–123.

Homem, C. C. F. and Knoblich, J. A. (2012). Drosophila neuroblasts: a model for stem cell biology. *Development* **139**, 4297–4310.

Homem, C. C. F., Repic, M. and Knoblich, J. A. (2015). Proliferation control in neural stem and progenitor cells. 1–13.

Hori, K., Sen, A. and Artavanis-Tsakonas, S. (2013). Notch signaling at a glance. *Journal of Cell Science* **126**, 2135–2140.

Housden, B. E., Fu, A. Q., Krejci, A., Bernard, F., Fischer, B., Tavaré, S., Russell, S. and Bray, S. J. (2013). Transcriptional Dynamics Elicited by a Short Pulse of Notch Activation Involves Feed-Forward Regulation by E(spl)/Hes Genes. *PLoS Genetics* **9**, e1003162.

Hurlbut, G. D., Kankel, M. W. and Artavanis-Tsakonas, S. (2009). Nodal points and complexity of Notch-Ras signal integration. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 2218–2223.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821.

Jovtchev, G., Borisova, B. E., Kuhlmann, M., Fuchs, J., Watanabe, K., Schubert, I. and Mette, M. F. (2011). Pairing of lacO tandem repeats in Arabidopsis thaliana nuclei requires the presence of hypermethylated, large arrays at two chromosomal positions, but does not depend on H3-lysine-9-dimethylation. *Chromosoma* **120**, 609–619.

Kalverda, B., Pickersgill, H., Shloma, V. V. and Fornerod, M. (2010). Nucleoporins Directly Stimulate Expression of Developmental and Cell-Cycle Genes Inside the Nucleoplasm. *Cell* **140**, 360–371.

Kind, J. and van Steensel, B. (2014). Stochastic genome-nuclear lamina interactions: modulating roles of Lamin A and BAF. *nucleus* **5**, 124–130.

Kind, J., Pagie, L., Ortabozkoyun, H., Boyle, S., de Vries, S. S., Janssen, H., Amendola, M., Nolen, L. D., Bickmore, W. A. and van Steensel, B. (2013). Single-Cell Dynamicsof Genome-Nuclear Lamina Interactions. *Cell* **153**, 178–192.

Kinney, N. A., Onufriev, A. V. and Sharakhov, I. V. (2015). Quantified effects of chromosome-nuclear envelope attachments on 3D organization of chromosomes. *nucleus* **6**, 212–224.

Knoblich, C. B. H. H. T. B. J. S. S. V. A.-S. L. M. N. H. R. J., Harzer, H., Burkard, T. R., Steinmann, J., van der Horst, S., Laurenson, A.-S., Novatchkova, M., Reichert, H. and Knoblich, J. A. (2012). FACS Purification and Transcriptome Analysis of Drosophila Neural Stem Cells Reveals a Role for Klumpfuss in Self-Renewal. *CellReports* **2**, 407–418.

Kobayashi, T. and Kageyama, R. (2014). *Expression Dynamics and Functions of Hes Factors in Development and Diseases*. 1st ed. Elsevier Inc.

Kohwi, M., Lupton, J. R., Lai, S.-L., Miller, M. R. and Doe, C. Q. (2013). Developmentally Regulated Subnuclear Genome Reorganization Restricts Neural Progenitor Competence in Drosophila. *Cell* **152**, 97–108.

Krejci, A., Bernard, F., Housden, B. E., Collins, S. and Bray, S. J. (2009). Direct Response to Notch Activation: Signaling Crosstalk and Incoherent Logic. *Sci Signal* **2**, ra1–ra1.

Langer-Safer, P. R., Levine, M. and Ward, D. C. (1982). Immunological method for mapping genes on Drosophila polytene chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4381–4385.

Larracuente, A. M. and Ferree, P. M. (2015). Simple Method for Fluorescence DNA In Situ Hybridization to Squashed Chromosomes. *JoVE*.

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., et al. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**, 289–293.

Mathog, D. and Sedat, J. W. (1989). The three-dimensional organization of polytene nuclei in male Drosophila melanogaster with compound XY or ring X chromosomes. *Genetics* **121**, 293–311.

Matzke, A. J. M. (2005). Use of Two-Color Fluorescence-Tagged Transgenes to Study Interphase Chromosomes in Living Plants. *PLANT PHYSIOLOGY* **139**, 1586–1596.

Mavromatakis, Y. E. and Tomlinson, A. (2016). R7 Photoreceptor Specification in the Developing Drosophila Eye: The Role of the Transcription Factor Deadpan. *PLoS Genetics* **12**, e1006159.

Mendjan, S., Taipale, M., Kind, J., Holz, H., Gebhardt, P., Schelder, M., Vermeulen, M., Buscaino, A., Duncan, K., Mueller, J., et al. (2006). Nuclear Pore Components Are Involved in the Transcriptional Regulation of Dosage Compensation in Drosophila. *Molecular Cell* **21**, 811–823.

Monastirioti, M., Giagtzoglou, N., Koumbanakis, K. A., Zacharioudaki, E., Deligiannaki, M., Wech, I., Almeida, M., Preiss, A., Bray, S. and Delidakis, C. (2010). Drosophila Hey is a target of Notch in asymmetric divisions during embryonic and larval neurogenesis. *Development* **137**, 191–201.

Mortin, L. I. and Sedat, J. W. (1982). Structure of Drosophila polytene chromosomes. Evidence for a toroidal organization of the bands. *Journal of Cell Science* **57**, 73–113.

Nowell, C. S. and Radtke, F. (2017). Notch as a tumour suppressor. *Nat Rev Cancer* **17**, 145–159.

Pascual-Garcia, P., Debo, B., Aleman, J. R., Talamas, J. A., Lan, Y., Nguyen, N. H., Won, K. J. and Capelson, M. (2017). Metazoan Nuclear Pores Provide a Scaffold for Poised Genes and Mediate Induced Enhancer- Promoter Contacts. *Molecular Cell* **66**, 63–76.e6.

Pecinka, A., Schubert, V., Meister, A., Kreth, G., Klatte, M., Lysak, M. A., Fuchs, J. R. and Schubert, I. (2004). Chromosome territory arrangement and homologous pairing in nuclei of Arabidopsis thaliana are predominantly random except for NOR-bearing chromosomes. *Chromosoma* **113**, 258–269.

Pickersgill, H., Kalverda, B., de Wit, E., Talhout, W., Fornerod, M. and van Steensel, B. (2006). Characterization of the Drosophila melanogaster genome at the nuclear lamina. *Nat Genet* **38**, 1005–1014.

Port, F., Chen, H.-M., Lee, T. and Bullock, S. L. (2014). Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. *Proc. Natl. Acad. Sci. U.S.A.*

Presente, A., Shaw, S., Nye, J. S. and Andres, A. J. (2002). Transgene-mediated RNA interference defines a novel role for notch in chemosensory startle behavior. *Genesis* **34**, 165–169.

Ragoczy, T., Telling, A., Scalzo, D., Kooperberg, C. and Groudine, M. (2014). Functional redundancy in the nuclear compartmentalization of the late-replicating genome. *nucleus* **5**, 626–635.

Raices, M. and D'Angelo, M. A. (2012). Nuclear pore complex composition: a new regulator of tissue-specific and developmental functions. *Nat. Rev. Mol. Cell Biol.* **13**, 687–699.

Ren, X., Sun, J., Housden, B. E., Hu, Y., Roesel, C., Lin, S., Liu, L.-P., Yang, Z., Mao, D., Sun, L., et al. (2013). Optimized gene editing technology for Drosophila melanogaster using germ line-specific Cas9. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19012–19017.

Ren, X., Yang, Z., Xu, J., Sun, J., Mao, D., Hu, Y., Yang, S.-J., Qiao, H.-H., Wang, X., Hu, Q., et al. (2014). Enhanced Specificity and Efficiency of the CRISPR/Cas9 System with Optimized sgRNA Parameters in Drosophila. *CellReports* **9**, 1151–1162.

Rieder, D., Trajanoski, Z. and McNally, J. G. (2012). Transcription factories. *Front Genet* **3**, 221.

Rohner, S., Kalck, V., Wang, X., Ikegami, K., Lieb, J. D., Gasser, S. M. and Meister, P. (2013). Promoter- and RNA polymerase II-dependent hsp-16 gene association with nuclear pores in Caenorhabditis elegans. *The Journal of Cell Biology* **200**, 589–604.

Roukos, V., Burgess, R. C. and Misteli, T. (2014). Generation of cell-based systems to visualize chromosome damage and translocations in living cells. *Nat Protoc* **9**, 2476–2492.

Rowley, M. J. and Corces, V. G. (2016). ScienceDirectThe three-dimensional genome: principles and roles of long-distance interactions. *Current Opinion in Cell Biology* **40**, 8–14.

Saad, H., Gallardo, F., Dalvai, M., Tanguy-le-Gac, N., Lane, D. and Bystricky, K. (2014). DNA Dynamics during Early Double-Strand Break Processing Revealed by Non-Intrusive Imaging of Living Cells. *PLoS Genetics* **10**, e1004187.

San-Juán, B. P. and Baonza, A. (2011). Developmental Biology. *Developmental Biology* **352**, 70–82.

Sanchez-Martin, M. and Ferrando, A. (2017). The NOTCH1-MYC highway toward T-cell acute lymphoblastic leukemia. *Blood* **129**, 1124–1133.

Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N. F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J. A., Umlauf, D., Dimitrova, D. S., et al. (2009). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet* **42**, 53–61.

Schwartz, Y. B. and Cavalli, G. (2017). Three-Dimensional Genome Organization and Function in Drosophila. *Genetics* **205**, 5–24.

Sexton, T., Kurukuti, S., Mitchell, J. A., Umlauf, D., Nagano, T. and Fraser, P. (2012). Sensitive detection of chromatin coassociations using enhanced chromosome conformation capture on chip. *Nat Protoc* **7**, 1335–1350.

Skalska, L., Stojnic, R., Li, J., Fischer, B., Cerda-Moya, G., Sakai, H., Tajbakhsh, S., Russell, S., Adryan, B. and Bray, S. J. (2015). Chromatin signatures at Notch-regulated enhancers reveal large-scale changes in H3K56ac upon activation. *EMBO J.* **34**, 1889–1904.

Skeath, J. B. and Doe, C. Q. (1998). Sanpodo and Notch act in opposition to Numb to distinguish sibling neuron fates in the Drosophila CNS. *Development* **125**, 1857–1865.

Song, F. and Stieger, K. (2017). Optimizing the DNA Donor Templatefor Homology-

Directed Repairof Double-Strand Breaks. *Mol Ther Nucleic Acids* **7**, 53–60.

Straight, A. F., Belmont, A. S., Robinett, C. C. and Murray, A. W. (1996). GFP tagging of budding yeast chromosomes reveals that protein-protein interactions can mediate sister chromatid cohesion. *Current Biology* **6**, 1599–1608.

Sun, H., Ghaffari, S. and Taneja, R. (2007). bHLH-Orange Transcription Factors in Development and Cancer. *Transl Oncogenomics* **2**, 107–120.

Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. and Vale, R. D. (2014). A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. *Cell* **159**, 635–646.

Terriente-Felix, A., Li, J., Collins, S., Mulligan, A., Reekie, I., Bernard, F., Krejci, A. and Bray, S. (2013). Notch cooperates with Lozenge/Runx to lock haemocytes into a differentiation programme. *Development* **140**, 926–937.

Ulvklo, C., MacDonald, R., Bivik, C., Baumgardt, M., Karlsson, D. and Thor, S. (2012). Control of neuronal cell fate and number by integration of distinct daughter cell proliferation modes with temporal progression. *Development* **139**, 678–689.

Urata, Y., Parmelee, S. J., Agard, D. A. and Sedat, J. W. (1995). A three-dimensional structural dissection of Drosophila polytene chromosomes. *The Journal of Cell Biology* **131**, 279–295.

van Steensel, B. and Belmont, A. S. (2017). Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169**, 780–791.

Vatolina, T. Y., Boldyreva, L. V., Demakova, O. V., Demakov, S. A., Kokoza, E. B., Semeshin, V. F., Babenko, V. N., Goncharov, F. P., Belyaeva, E. S. and Zhimulev, I. F. (2011). Identical Functional Organization of Nonpolytene and Polytene Chromosomes in Drosophila melanogaster. *PLoS ONE* **6**, e25960.

Wang, C., Yao, C., Li, Y., Cai, W., Bao, X., Girton, J., Johansen, J. and Johansen, K. M. (2013). Evidence against a Role for the JIL-1 Kinase in H3S28 Phosphorylation and 14-3-3 Recruitment to Active Genes in Drosophila. *PLoS ONE* **8**, e62484.

Wang, R., Kamgoue, A., Normand, C., Léger-Silvestre, I., Mangeat, T. and Gadal, O. (2016a). High resolution microscopy reveals the nuclear shape of budding yeast during cell cycle and in various biological states. *Journal of Cell Science* **129**, 4480–4495.

Wang, R., Normand, C. and Gadal, O. (2016b). High-Throughput Live-Cell Microscopy Analysis of Association Between Chromosome Domains and the Nucleolus in S. cerevisiae. *Methods Mol Biol* **1455**, 41–57.

Wendorff, A. A., Koch, U., Wunderlich, F. T., Wirth, S., Dubey, C., Brüning, J. C.,

MacDonald, H. R. and Radtke, F. (2010). Hes1 Is a Critical but Context-Dependent Mediator of Canonical Notch Signaling in Lymphocyte Development and Transformation. *Immunity* **33**, 671–684.

White, R. (2012). Packaging the fly genome: domains and dynamics. *Brief Funct Genomics* **11**, 347–355.

Williamson, I., Lettice, L. A., Hill, R. E. and Bickmore, W. A. (2016). Shhand ZRS enhancer colocalisation is specific to the zone of polarising activity. *Development* **143**, 2994–3001.

Yashiro-Ohtani, Y., Wang, H., Zang, C., Arnett, K. L., Bailis, W., Ho, Y., Knoechel, B., Lanauze, C., Louis, L., Forsyth, K. S., et al. (2014). Long-range enhancer activity determines Mycsensitivity to Notch inhibitors in T cell leukemia. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E4946–E4953.

Zacharioudaki, E., Magadi, S. S. and Delidakis, C. (2012). bHLH-O proteins are crucial for Drosophila neuroblast self-renewal and mediate Notch-induced overproliferation. *Development* **139**, 1258–1269.

Zhu, S., Wildonger, J., Barshow, S., Younger, S., Huang, Y. and Lee, T. (2012). The bHLH Repressor Deadpan Regulates the Self-renewal and Specification of Drosophila Larval Neural Stem Cells Independently of Notch. *PLoS ONE* **7**, e46724.

Zielke, T., Glotov, A. and Saumweber, H. (2016). High-resolution in situ hybridization analysis on the chromosomal interval 61C7-61C8 of Drosophila melanogaster reveals interbands as open chromatin domains. *Chromosoma* 1–13.

8 Appendices

APPENDIX 1: OLIGONUCLEOTIDES

Type	ID	Oligo name	Oligo sequence (5' to 3')
gRNA	M139	ebony - gRNA F	AAACTGACGATCGACAATTGTGGC
gRNA	M138	ebony - gRNA R	GTCGGCCACAATTGTCGATCGTCA
HDR	M171	ebony - HDR F 1	CATTGCGGCCGCGTCTGCAGCAAGACTTCG
HDR	M170	ebony - HDR R 1	ATGCGAATTCTCCGTGGATCCGTAAAAGTTG
HDR	M173	ebony - HDR F 2	ATATGCTCTTCTGACCCGCATTCCGAGCGCTTG
HDR	M172	ebony - HDR R 2	ATATGCTCTTCGTATCGATCGACAATTGTGG
gRNA	M165	E(spl)-mδ - gRNA F	GTCGCCTGCCGGGTCGGGGGGATC
gRNA	M164	E(spl)-mδ - gRNA R	AAACGATCCCCCCGACCCGGCAGG
HDR	M262	E(spl)-mδ - HDR F 1	GGTCCTGCCGGGTCGGGGGGGGACATATGCACACCTGCGA
HDR	M261	E(spl)-mδ - HDR R 1	TCGCAGGTGTGCATATGTCCCCCCCCGACCCGGCAGGACC
HDR	M260	E(spl)-mδ - HDR F 2	GCTTAATGCGTGTACGACTTTGCTCTTCTATAACTTCGTATAGC
HDR	M259	E(spl)-mδ - HDR R 2	TACGAAGTTATAGAAGAGCAAAGTCGTACACGCATTAAGC

Type	ID	Oligo name	Oligo sequence (5' to 3')
HDR	M205	E(spl)-mδ - HDR F 3	AACTCGATTGACGGAAGAGCCTTTCTCCAAACGGGTGACGTGCTCG
HDR	M204	E(spl)-mδ - HDR R 3	TACGAAGTTATAGAAGAGCAGCCGGGTCGGGGGGATCGGG
HDR	M203	E(spl)-mδ - HDR F 4	CCCGATCCCCCCGACCCGGCTGCTCTTCTATAACTTCGTATAGC
HDR	M202	E(spl)-mδ - HDR R 4	CGGATGGAGGGAGAGGCAGAGGACATATGCACACCTGCGA
HDR	M201	E(spl)-mδ - HDR F 5	TCGCAGGTGTGCATATGTCCTCTGCCTCTCCCTCCATCCG
HDR	M200	E(spl)-mδ - HDR R 5	CGCTGAAGCAGGTGGAATTCTCACCAGCTCAAGGACATGA
HDR	M199	E(spl)-mδ - HDR F 6	TCATGTCCTTGAGCTGGTGAGAATTCCACCTGCTTCAGCG
HDR	M198	E(spl)-mδ - HDR R 6	CGTCACCCGTTTGGAGAAAGGCTCTTCCGTCAATCGAGTT
gRNA	M277	E(spl)-m8 - gRNA F	AAACTTACAAATCTTGAGGGTTCT
gRNA	M276	E(spl)-m8 - gRNA R	GTCGAGAACCCTCAAGATTTGTAA
HDR	M274	E(spl)-m8 - HDR F 1	TGAGATGGATTTGTATATCTTGCTCTTCTATAACTTCGTATAGC
HDR	M273	E(spl)-m8 - HDR R 1	ACAGACTTAACTTAGAAACCAGAATTCCACCTGCTTCAGCG
HDR	M272	E(spl)-m8 - HDR F 2	AGAATATCACTCACGATTACAGGCTCTTCCGTCAATCGAGTT
HDR	M271	E(spl)-m8 - HDR R 2	CTCGATTGACGGAAGAGCCTGTAATCGTGAGTGATATTCTG

Type	ID	Oligo name	Oligo sequence (5' to 3')
HDR	M270	E(spl)-m8 - HDR F 3	TACGAAGTTATAGAAGAGCAAGATATACAAATCCATCTCAATT
HDR	M269	E(spl)-m8 - HDR R 3	TTAAGAACCCTCAAGATTTGGGACATATGCACACCTGCGA
HDR	M268	E(spl)-m8 - HDR F 4	TCGCAGGTGTGCATATGTCCCAAATCTTGAGGGTTCTTAATATTG
HDR	M267	E(spl)-m8 - HDR R 4	GCTGAAGCAGGTGGAATTCTGGTTTCTAAGTTAAGTCTGTGATAG
gRNA	M331	*dpn* - gRNA F	AAACCTCGTCGCTAAGCGCCATAA
gRNA	M330	*dpn* - gRNA R	GTCGTTATGGCGCTTAGCGACGAG
HDR	M332	*dpn* - HDR F 1	GCTGAAGCAGGTGGAATTCTGCAAATCTTTACAGCGAACAAATCC
HDR	M341	*dpn* - HDR R 1	GATTTGTTCGCTGTAAAGATTTGCAGAATTCCACCTGCTTCAGCG
HDR	M360	*dpn* - HDR F 2	CGCAGGTGTGCATATGTCCCAAACCTGAGTGAAGTGTGCG
HDR	M361	*dpn* - HDR R 2	GCACACTTCACTCAGGTTTGGGACATATGCACACCTGCGA
HDR	M338	*dpn* - HDR F 3	TACGAAGTTATAGAAGAGCAGTCGCTAAGCGCCATAAAAATC
HDR	M337	*dpn* - HDR R 3	TTTTTATGGCGCTTAGCGACTGCTCTTCTATAACTTCGTATAGC
HDR	M339	*dpn* - HDR F 4	CTCGATTGACGGAAGAGCCTATCATTACGGAGCTTACCTGGC
HDR	M340	*dpn* - HDR R 4	CAGGTAAGCTCCGTAATGATAGGCTCTTCCGTCAATCGAGTT

Type	ID	Oligo name	Oligo sequence (5' to 3')
gRNA	M319	Hey - gRNA F	AAACCGCAGTGACAAAGCGCACCT
gRNA	M318	Hey - gRNA R	GTCGAGGTGCGCTTTGTCACTGCG
HDR	M320	Hey - HDR F 1	CCCTTCGCTGAAGCAGGTGGCCCTACTTGTCAGGTGGCAG
HDR	M329	Hey - HDR R 1	CTGCCACCTGACAAGTAGGGAGAATTCCACCTGCTTCAGCG
HDR	M362	Hey - HDR F 2	TATGACGGCGAACTGGTAGTGGACATATGCACACCTGCGA
HDR	M363	Hey - HDR R 2	TCGCAGGTGTGCATATGTCCACTACCAGTTCGCCGTCATAG
HDR	M326	Hey - HDR F 3	TACGAAGTTATAGAAGAGCAAGTGACAAAGCGCACCTCAAG
HDR	M325	Hey - HDR R 3	GGTGCGCTTTGTCACTTGCTCTTCTATAACTTCGTATAGC
HDR	M328	Hey - HDR F 4	GGGCAGCGATTCAGTTACAGGCTCTTCCGTCAATCGAGTT
HDR	M327	Hey - HDR R 4	CTCGATTGACGGAAGAGCCTGTAACTGAATCGCTGCCCAGC
genotyping	M112	m3 geno F 1	CACATTCACACGGTACCCAG
genotyping	M120	m3 geno R 1	CATTAGATCCCCCATACCGATG
genotyping	M176	ebony geno F 1	GATCAGAGCCACCTTGTCG
genotyping	M175	ebony geno R 1	GCCAGCTCCCATATTCATACTC

Type	**ID**	**Oligo name**	**Oligo sequence (5' to 3')**
genotyping	M275	DsRed geno R 1	TGAAGCGCATGAACTCCTTG
genotyping	M174	DsRed geno F 1	CCACCACCTGTTCCTGTAG
genotyping	M307	ebony.attP geno R	CGGGTAATTGTCGTTCTTGG
genotyping	M313	mdelta geno R	AAGACTTCCTTTTCTGCTGG
genotyping	M309	m8 geno R	GGGTCTTGGTGGTGTATTCC
genotyping	M312	mdelta geno F	GGAGAAAGCTGACATCTTAGAG
genotyping	M308	m8 geno F	AAGGACTGAATTCCCGACG
genotyping	M226	mdelta geno F 2	GAACTCTCCCCACTCCGTTTAC
genotyping	M377	attB geno 1	CGACCCGTTCATCATGATG
genotyping	M316	m8 geno F 2	GTCCCTGGTAAACAAAGATCTAC
sequencing	M374	Hey seq 1	TGCTGATTAGACCCGGATTAC
sequencing	M376	*Dpn* seq 1	CGACAGATGTTAGGTCTTGAAATG
qPCR		m3 real sense	AGCCCACCCACCTCAACCAG
qPCR		m3 real anti	CGTCTGCAGCTCAATTAGTC

Type	ID	Oligo name	Oligo sequence (5' to 3')
qPCR		m8.1	CCAGAAAGCAATGCGAGCAG
qPCR		m8.2	TTCGTAGGACGGAGGACAATCC
qPCR		m7 sense	CGTTGCTCAGACTGGCGATG
qPCR		m7 anti	CTCGTTGTCGCTGGCATATC
qPCR		mctrl short for	CAATTCCACGAAGCACAGTC
qPCR		mctrl short rev	GAGGAGCAGTCCATCGAGTT
qPCR		mctrl short for	CAATTCCACGAAGCACAGTC
qPCR		mctrl short rev	GAGGAGCAGTCCATCGAGTT
qPCR		mbe real	AGAAGTGAGCAGCAGCCATC
qPCR		mbe real	GCTGGACTTGAAACCGCACC
qPCR		m7 sense	CGTTGCTCAGACTGGCGATG
qPCR		m7 anti	CTCGTTGTCGCTGGCATATC
qPCR		m3 real sense	AGCCCACCCACCTCAACCAG

APPENDIX 2: GENOTYPING DATA FOR *DPN* AND *HEY* ATTP SITES

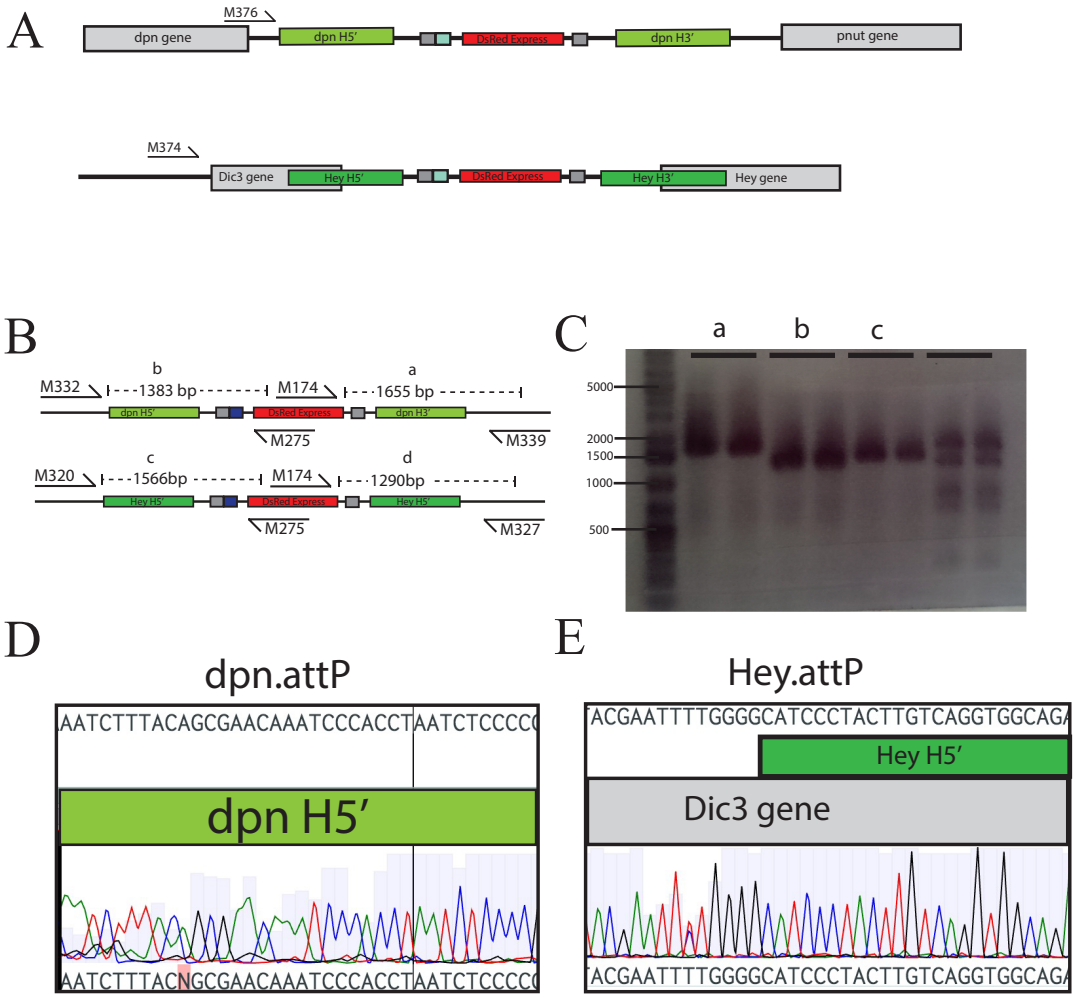


Figure 8.1: Confirmation of pHD-DsRed.attP integration into intergenic regions adjacent to *dpn* and *Hey*.

Schematic of homology arms relative to target loci and neighbouring genes (A). Diagram of PCR amplicons used to detect successful integration into the desired locus in the correct orientation (B). Agarose gel showing amplicons shown in B, (C). Sequence trace from primer annealing in genomic sequence not contained within the repair template plasmid showing *dpn* 5' homology arm is fused to intergenic sequence (Primer M332 as shown in B) (D). Sequence trace showing junction between Hey 5' homology arm and Dic3 sequence not present in the repair template plasmid (Primer M320) (E).

APPENDIX 3: FLYSTOCKS

Name	**Genotype**	**Purpose**	**Bloomington Stock number**	**Origin**
nos-Cas9/pCFD2	y1 P(nos-cas9, w+) M(3xP3-RFP.attP)ZH-2A w*	germline expression of Cas9	NA	CRISPR Fly Design/Simon Bullock Lab
yellow-white	yw;;	wild type control	NA	Bray Lab
TM3/TM6B	w-;;TM3/TM6B	Third Chromosome Balancer/Ebony screening	NA	Bray Lab
double balancer	w-;Sco/CyoYFP;Dr/TM6B	maintain transgenes on 2nd and 3rd chromosome	NA	Bray Lab
34770	y[1] w[*] P{y[+t7.7]=nos-phiC31int.NLS}X; sna[Sco]/CyO	ΦC31 integrase expression from the X, balancers on 2nd	34470	Bloomington
34771	y[1] w[*] P{y[+t7.7]=nos-phiC31int.NLS}X; Dr[1] e[1]/TM3, Sb[1]	ΦC31 integrase expression from the X, balancers on 3rd	34471	Bloomington

Name	Genotype	Purpose	Bloomington Stock number	Origin
ParB1-mCherry	10967-2 pAttb-UAST-p31A-mCherry 24749 86F8	Express Locus tag protein	NA	Payre Lab
ParB2-mCherry	10967-4 pAttb-UAST-p31BmCherrry 24749 86F8	Express Locus tag protein	NA	Payre Lab
ParB1-GFP	10967-1 pAttb-UAST-p31A-GFP 24749 86F8	Express Locus tag protein	NA	Payre Lab
ParB2-GFP	10967-3 pAttb-UAST-p31B-GFP 24749 86F8	Express Locus tag protein	NA	Payre Lab
Ctrl.86Fb.Int1	10967-5 pAttb-IntA-hsLacZ 24749 (86F8)	Control Locus tag	NA	Payre Lab
Ctrl.68E1.Int1	11872-5 pAttb-IntA-hsLacZ 24485 (68E1)	Control Locus tag	NA	Payre Lab
Ctrl.86Fb.Int2	10967-6 pAttb-IntB-	Control Locus tag	NA	Payre Lab

Name	Genotype	Purpose	Bloomington Stock number	Origin
	hsLacZ 24749 (86F8)			
Ctrl.68E1.Int2	11872-6 pAttb-IntB-hsLacZ 24485 (68E1)	Control Locus tag	NA	Payre Lab
E(spl)-m8.Int1	w-;; P(pAttb-IntA-hsLacZ) E(spl)-m8.attP	E(sp) locus tag	NA	This study
E(spl)-m8.Int2	w-;; P(pAttb-IntB-hsLacZ) E(spl)-m8.attP	E(sp) locus tag	NA	This study
E(spl)-mdelta.Int1	w-;; P(pAttb-IntA-hsLacZ) E(spl)-mdelta.attP	E(sp) locus tag	NA	This study
E(spl)-mdelta.Int2	w-;; P(pAttb-IntB-hsLacZ) E(spl)-mdelta.attP	E(sp) locus tag	NA	This study
Hey.Int1	w-; P(pAttb-IntA-hsLacZ) Hey.attP;	Hey locus tag	NA	This study
dpn.Int2	w-; P(pAttb-IntB-hsLacZ) dpn.attP;	dpn locus tag	NA	This study

Name	**Genotype**	**Purpose**	**Bloomington Stock number**	**Origin**
Notch-RNAi	[P{w[+mC]=UAS-N.dsRNA.P}14E, w[*]].	Notch knockdown	7078	Bloomington
NΔECD	;UAS-NΔECD/Cyo;Dr/Tm6B	Notch overexpression	NA	Artavanis Lab
1151-Gal4	1151-Gal4;;TM2/TM6B	Drive UAS construct expression in salivary gland	NA	L. S. Shashidhara
Inscutable-Gal4 ~ Gal80ts	w*; P{GawB}inscMz1407 ~ Gal80ts	Drive UAS construct expression in neuroblast lineages	8751	Bloomington
E(spl)-m8.Int1 ~ E(spl)-mdelta.Int2	w-;;P(pAttb-IntB-hsLacZ) E(spl)-mdelta.attP ~ P(pAttb-IntB-hsLacZ) E(spl)-m8.attP	Dual E(spl) locus tags	NA	This study (made by ZP and MGL)
ParB1-mCherry ~ ParB2-	P(p31A-mCherry)	Dual Par expression	NA	This study (made by ZP

Name	**Genotype**	**Purpose**	**Bloomington Stock number**	**Origin**
GFP	68E1.attP ~ P(p31B-GFP) 86Fb.attP			and MGL)

APPENDIX 4: NEUROBLAST CULTURE PROTOCOL

S+G medium

Supplement the Schneider medium (GIBCO, ref number 21720-024) with 1mg/ml Glucose (D-L-glucose monohidrate).

Filter-sterilize it and keep at 7ºC.

Culture Medium

(Prepare fresh before each culture)

Add to the S+G medium:

10% Fetal Calf Serum

2.5% fly extract (see below the recipe)

10 μg/ml human Insulin (Sigma, 19278 – 5mL)

1x Pen/Strep (ref: 15140, From Gibco)

Fly extract (according to Milner's lab protocol)

Collect adult flies in a 50 ml plastic disposable centrifuge tube and place in a 20ºC freezer for at least 45 minutes.

Weigh 2gr of the flies and transfer them into a glass homogenizer, together with 6.8 ml of S+G medium/gr of flies. Homogenize on ice.

Spin the homogenate at 1500X g at 4ºC for 15 minutes. Decant the supernatant into fresh tubes. Discard the pellet.

Incubate the supernatant at 60ºC for 5 minutes to inactivate tyrosinase.

Spin at 70000 rpm at 4ºC for 20 minutes. Collect the supernatant. This is the fly extract.

Filter-sterilize the extract through a 0.22 μm filter.

Aliquot and place at -20ºC.

The extract is stable for years at -20ºC.

Preparation of Poly-D-Lysin coated plates

Pipet 800 μl of nitric acid onto glassbottom fluorodish

Leave them over night ca15h

Then, pipet off the acid and rinse 2x with sterile water.

Wash 3x, 2h each, with sterile water

Pipet 1ml of 1mg/ml polylysine solution onto the glass bottom dish.

Leave over night ca. 15h

Then, briefly rinse 2x with sterile water.

Wash 2x 1h with sterile water.

UV irradiate (under the hood) for 20 minutes and keep sterile.

10x Collagenase solution

Prepare 2mg/ml of collagenase (from Sigma, c0130) in Dissecting media (below the recipe), store at -20ºC.

Dissecting media=collagenase buffer (x10)

Weight:

8 g of NaCl

0.2 g of KCl

0.05 g of NaH_2PO_4

1 g of $NaHCO_3$

1 g of glucose

and dilute it in 100ml distilled water

Keep at 7ºC.

Neuroblast culture method

Dissect the brains in dissecting media 1x.

Transfer dissected brains to collagenase solution 1x (diluted in dissecting media) and incubate 15 minutes (this might vary depending on the collagenase batch. I never had to leave them for more than 30 mins).

Remove the collagenase 1x and rinse brains 3 times with culture medium

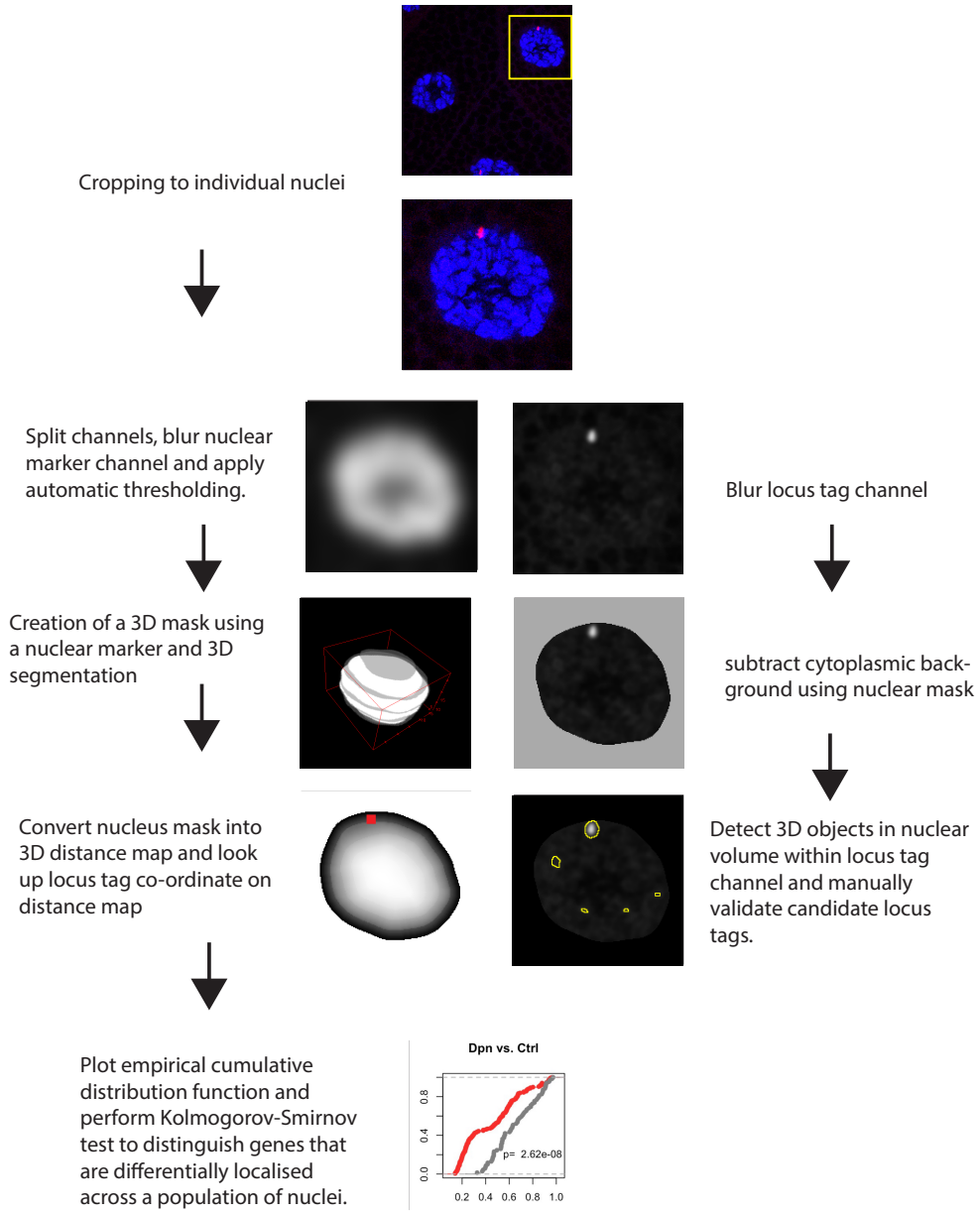
Transfer the brains together with 40μl of medium/brain into an eppendorf tube and dissociate into single cells by pipetting up and down.

Place the pipette-dissociated cells onto Poly-D-Lysine coated plates.

Let them rest for 30 mins before imaging.

APPENDIX 5: LOCUS TAG QUANTIFICATION SCHEMATIC

An open source image processing pipeline for reproducible analysis of nuclear morphology and locus tag position quantification



APPENDIX 6: P-VALUES FROM PAIR-WISE KOLMOGOROV-SMIRNOV TESTS ON SALIVARY GLAND LOCUS TAG DATA IN FIGURE 4.8

	Ctrl.68E1	Ctrl.86Fb	rand1
Ctrl.68E1	-	0.92	0.73
Ctrl.86Fb	0.92	-	0.43
rand1	0.73	0.43	-
	md	m8	rand2
E(spl)-mdelta	-	0.56	0.02
E(spl)-m8	0.56	-	0.01
rand2	0.02	0.01	-
	Hey	Dpn	rand3
Hey	-	0.0000001	0.0000179
dpn	0.0000001	-	0.0006509
rand3	0.0000179	0.0006509	-